PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : A01N 63/00 — A1

(11) International Publication Number: **WO 00/41566**

(43) International Publication Date: 20 July 2000 (20.07.00)

(21) International Application Number: PCT/US00/00409

(22) International Filing Date: 6 January 2000 (06.01.00)

(30) Priority Data:
09/229,037 12 January 1999 (12.01.99) US

(71) Applicant *(for all designated States except US)*: SANAGAMO BIOSCIENCES, INC. [US/US]; Suite A100, 501 Canal Boulevard, Richmond, CA 94804 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: COX, George, Norbert, III [US/US]; 678 West Willow Street, Louisville, CO 80027 (US). CASE, Casey, Christopher [US/US]; 605 Barroilhet Avenue, San Mateo, CA 94402 (US). EISENBERG, Stephen, P. [US/US]; 5664 Pennsylvania Place, Boulder, CO 80303 (US). JARVIS, Eric, Edward [US/US]; 3720 Smuggler Place, Boulder, CO 80303 (US). SPRATT, Sharon, Kaye [US/US]; 107 Edmondale Court, Vacaville, CA 95688 (US).

(74) Agents: PARENT, Annette, S. et al.; Townsend & Townsend & Crew LLP, 8th Floor, Two Embarcadero Center, San Francisco, CA 94111–3834 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: REGULATION OF ENDOGENOUS GENE EXPRESSION IN CELLS USING ZINC FINGER PROTEINS

(57) Abstract

The present invention provides methods for modulating expression of endogenous cellular genes using recombinant zinc finger proteins.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

BNSDOCID: <WO___0041566A1_I_>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/00409

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : A01N 63/00
US CL : 424/93.2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 424/93.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
CAPLUS, MEDLINE, SCISEARCH, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BEERLI et al. Toward controlling gene expression at will: Specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc. Natl. Acad. Sci. USA. December 1998, Vol.95, pages 14628-14633, see entire document.	1-90

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 March 2000 (24.03.2000)	04 APR 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Jeffrey S. Lundgren Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

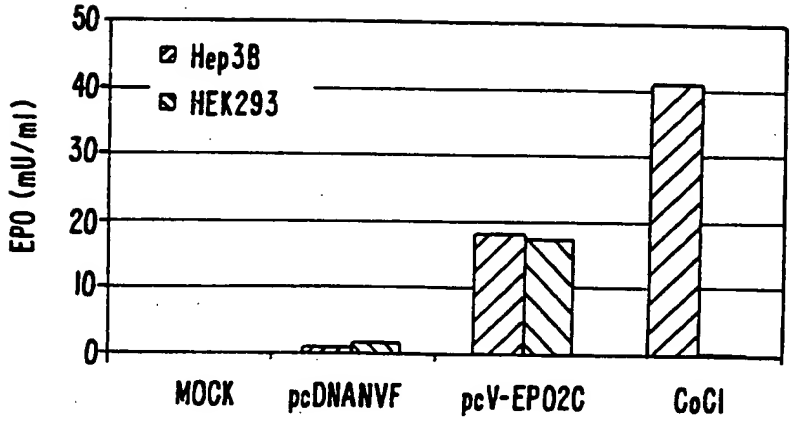


FIG. 11A.

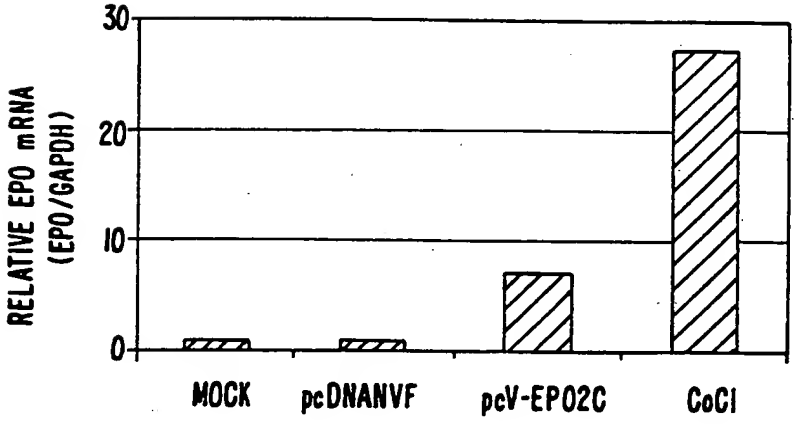


FIG. 11B.

FIG. 10B.

FIG. 10A.

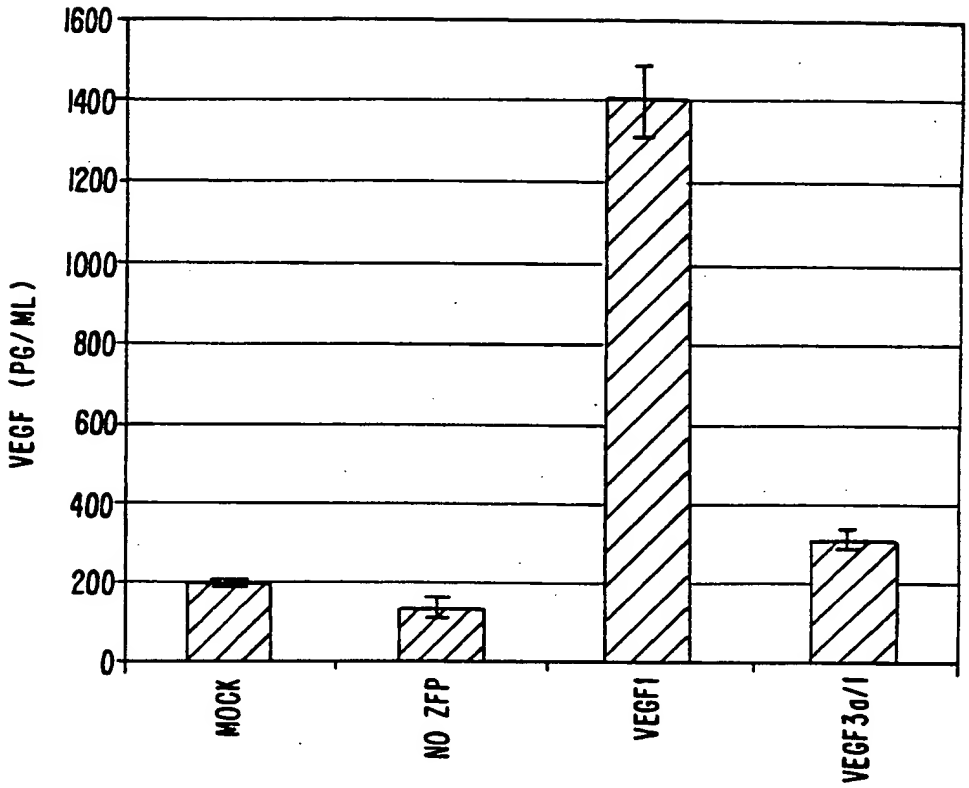

1600
1400
1200
1000
800
600
400
200
0
VEGF (PG/ML)
MOCK
NO ZFP
VEGF1
VEGF3d/I

FIG. 9.

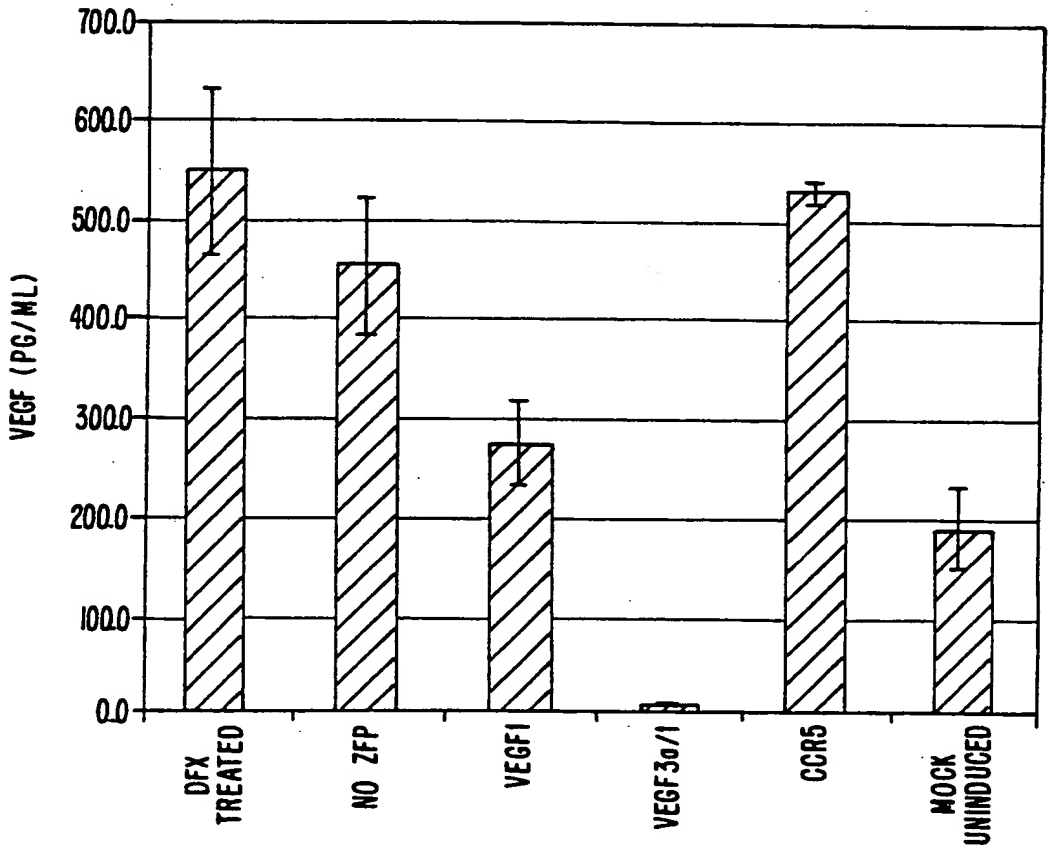

700.0
600.0
500.0
400.0
300.0
200.0
100.0
0.0
VEGF (PG/ML)
DFX TREATED
NO ZFP
VEGF1
VEGF3o/1
CCR5
MOCK UNINDUCED

FIG. 8.

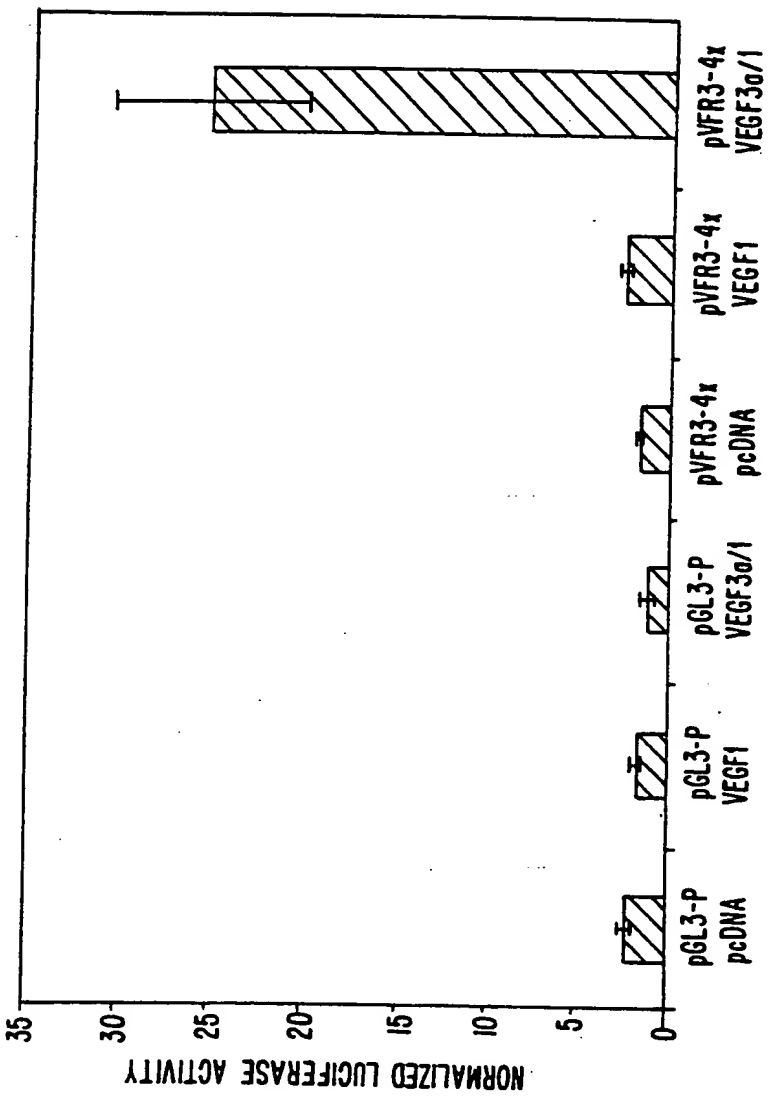

NORMALIZED LUCIFERASE ACTIVITY
0
5
10
15
20
25
30
35
pGL3-P pcDNA
pGL3-P VEGF1
pGL3-P VEGF3a/1
pVFR3-4x pcDNA
pVFR3-4x VEGF1
pVFR3-4x VEGF3a/1

FIG. 7.

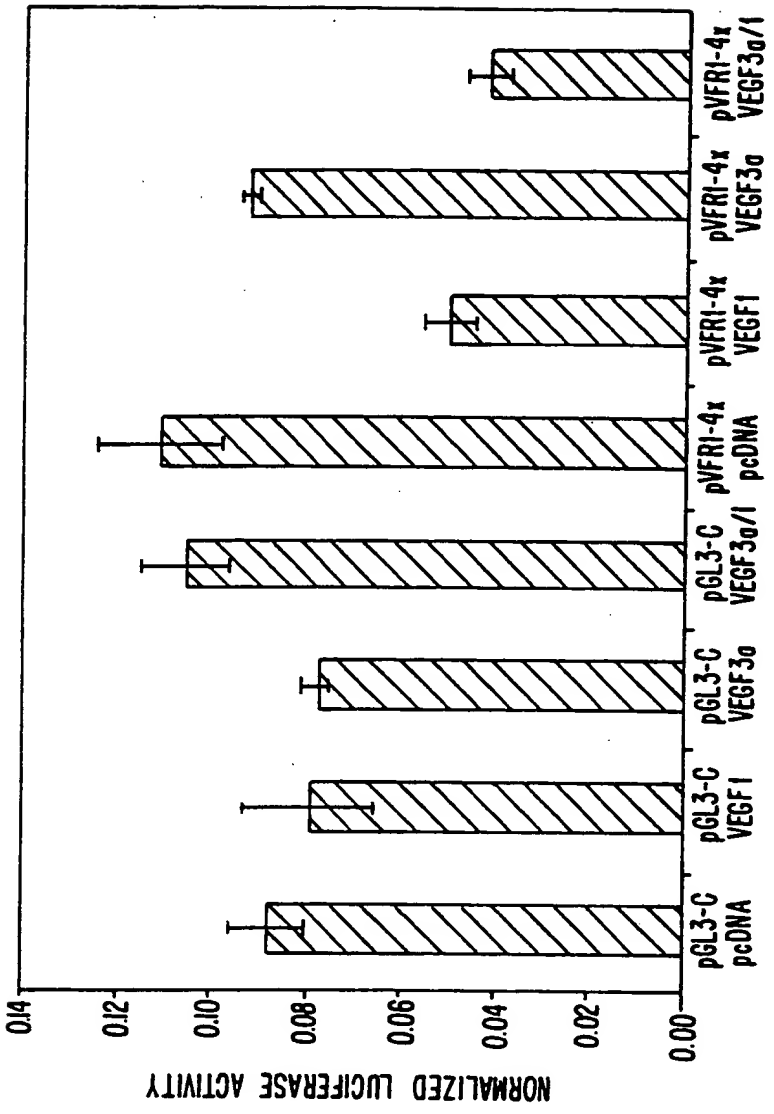

NORMALIZED LUCIFERASE ACTIVITY
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
pGL3-C pcDNA
pGL3-C VEGF1
pGL3-C VEGF3α
pGL3-C VEGF3α/1
pVFR1-4x pcDNA
pVFR1-4x VEGF1
pVFR1-4x VEGF3α
pVFR1-4x VEGF3α/1

FIG. 6.

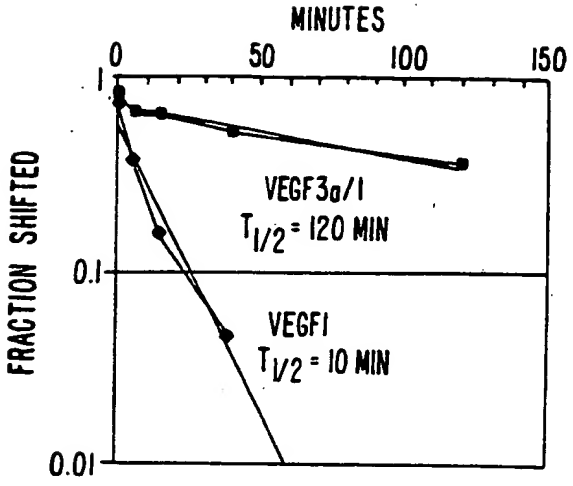

MINUTES
0
50
100
150
1
0.1
0.01
FRACTION SHIFTED
VEGF3a/1
$T_{1/2}$ = 120 MIN
VEGF1
$T_{1/2}$ = 10 MIN

FIG. 4B.

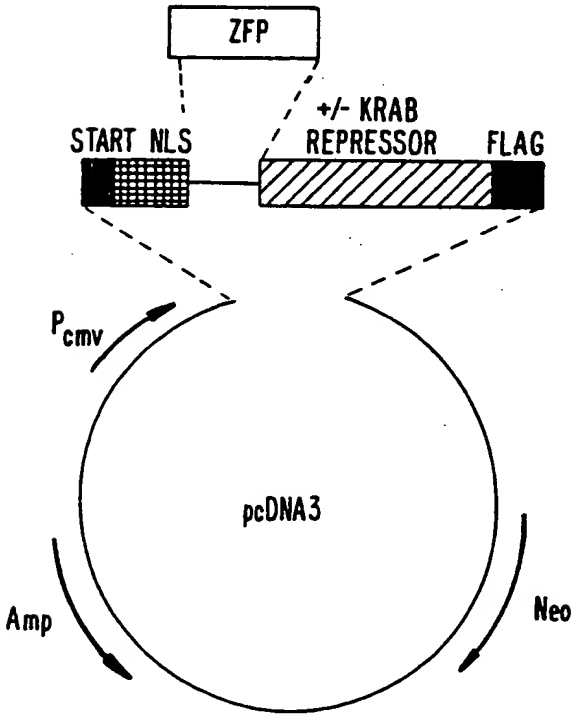
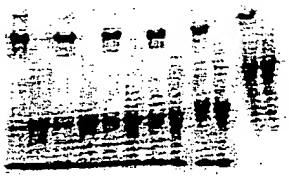

ZFP
+/- KRAB
START NLS
REPRESSOR
FLAG
P_{cmv}
pcDNA3
Amp
Neo

FIG. 5.

VEGF3a/1



VEGF1

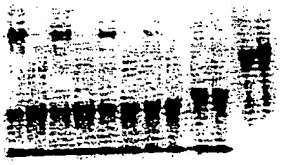


FIG. 4A.

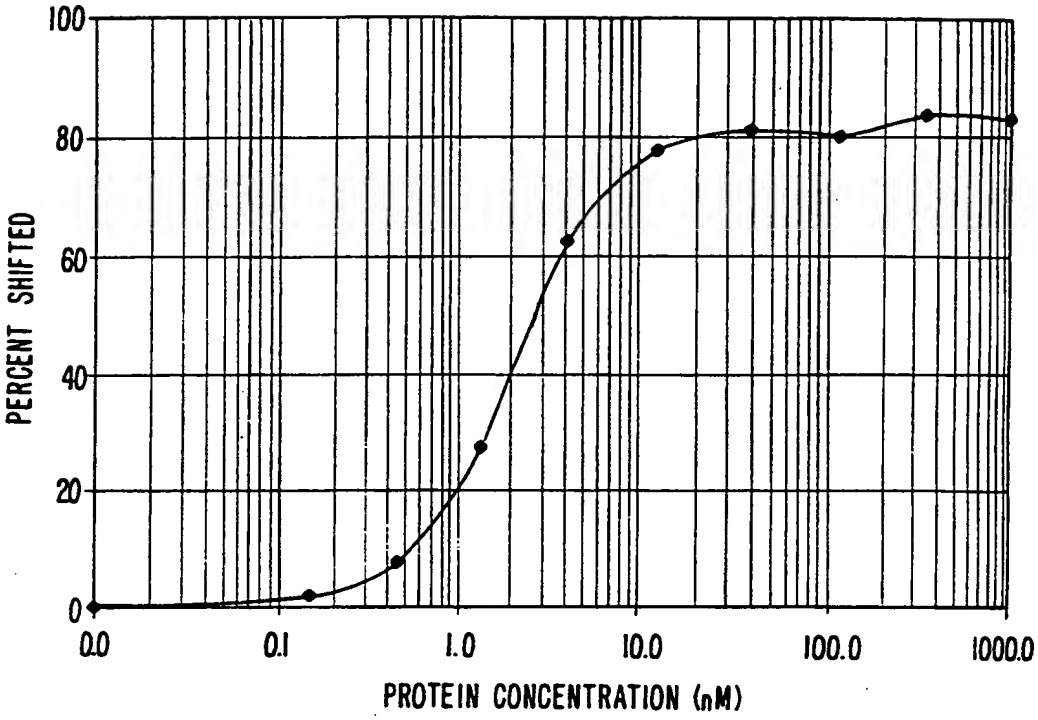


FIG. 3.

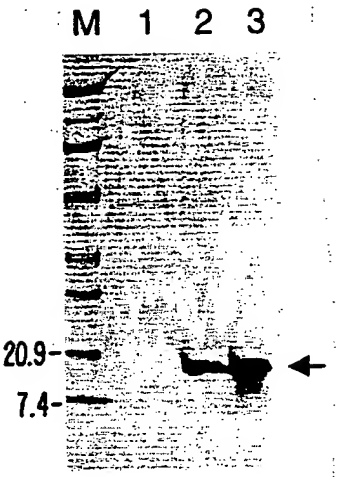


FIG. 2A.

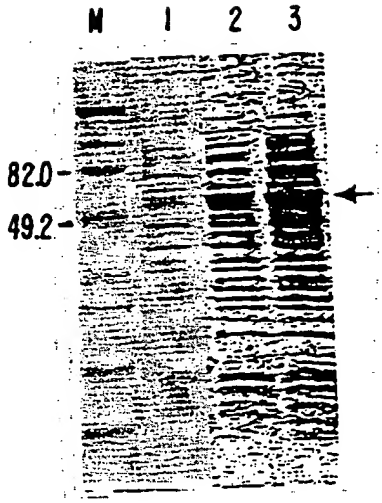


FIG. 2B.

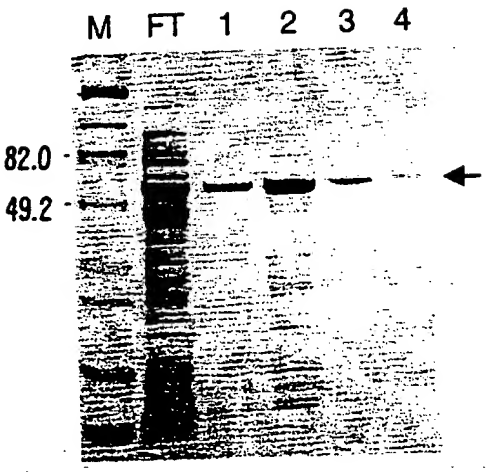


FIG. 2C.

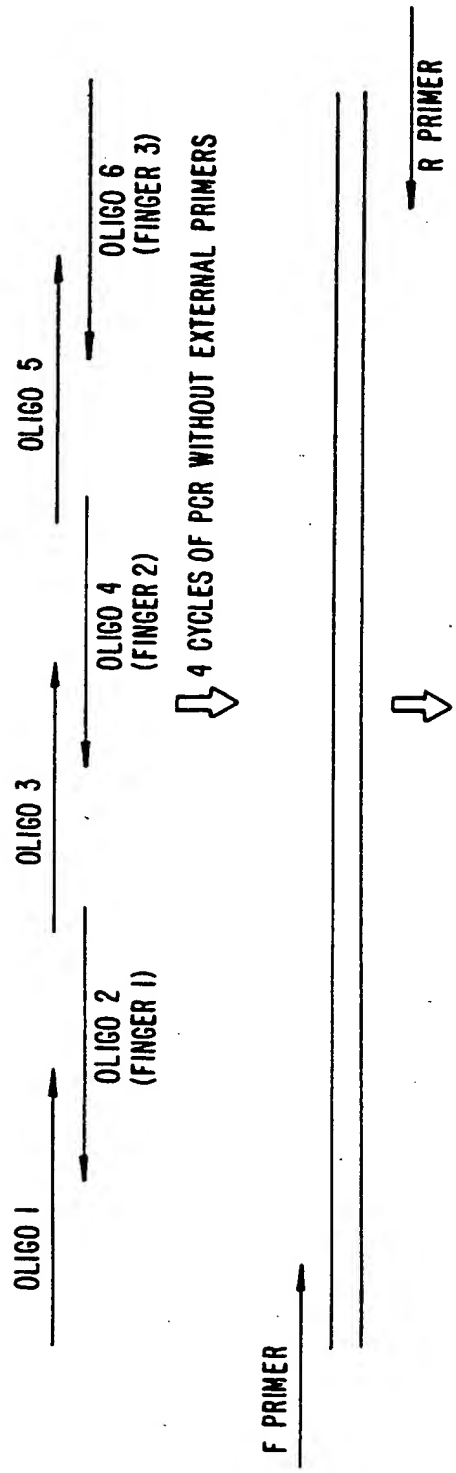


FIG. 1.

82. The method of claim 79, wherein the zinc finger protein is encoded by a nucleic acid operably linked to an inducible promoter.

83. The method of claim 79, wherein the zinc finger protein is encoded by a nucleic acid operably linked to a weak promoter.

84. The method of claim 63, wherein the cell comprises less than about $1.5x10^6$ copies of the zinc finger protein.

85. The method of claim 63, wherein the target site is upstream of a transcription initiation site of the endogenous cellular gene.

86. The method of claim 63, wherein the target site is adjacent to a transcription initiation site of the endogenous cellular gene.

87. The method of claim 63, wherein the target site is adjacent to an RNA polymerase pause site downstream of a transcription initiation site of the endogenous cellular gene.

88. The method of claim 63, wherein the zinc finger protein comprises an SP-1 backbone.

89. The method of claim 89, wherein the zinc finger protein comprises a regulatory domain and is humanized.

90. A method of modulating expression of an endogenous cellular gene in a cell, the method comprising the step of:

contacting a target site in the endogenous cellular gene with a fusion zinc finger protein comprising six fingers and a regulatory domain;

thereby modulating expression of the endogenous cellular gene.

72. The method of claim 71, wherein the cell is a mammalian cell

73. The method of claim 72, wherein the cell is a human cell.

74. The method of claim 63, wherein the endogenous cellular gene is a selected from the group consisting of VEGF, ERα, IGF-I, c-myc, c-myb, ICAM, Her2/Neu, FAD2-1, EPO, GM-CSF, GDNF, and LDL-R.

75. The method of claim 63, wherein the endogenous cellular gene is VEGF.

76. The method of claim 67 or 69, wherein the regulatory domain is selected from the group consisting of a transcriptional repressor, a transcriptional activator, an endonuclease, a methyl transferase, a histone acetyltransferase, and a histone deacetylase.

77. The method of claim 63, wherein the method further comprises the step of first administering to the cell a delivery vehicle comprising the zinc finger protein, wherein the delivery vehicle comprises a liposome or a membrane translocation polypeptide.

78. The method of claim 63, wherein the zinc finger protein is encoded by a zinc finger protein nucleic acid operably linked to a promoter, and wherein the method further comprises the step of first administering the nucleic acid to the cell in a lipid:nucleic acid complex or as naked nucleic acid.

79. The method of claim 63, wherein the zinc finger protein is encoded by an expression vector comprising a zinc finger protein nucleic acid operably linked to a promoter, and wherein the method further comprises the step of first administering the expression vector to the cell.

80. The method of claim 79, wherein the expression vector is a viral expression vector.

81. The method of claim 79, wherein the expression vector is a retroviral expression vector, an adenoviral expression vector, or an AAV expression vector.

contacting a target site in the endogenous cellular gene with a fusion zinc finger protein comprising six fingers and a regulatory domain, wherein the K_d of the zinc finger protein is less than about 25 nM;

thereby activating expression of the endogenous cellular gene to at least about 150%.

63. A method of modulating expression of an endogenous cellular gene in a cell, the method comprising the step of:

contacting a first target site in the endogenous cellular gene with a first zinc finger protein;

thereby modulating expression of the endogenous cellular gene.

64. The method of claim 63, wherein the step of contacting further comprises contacting a second target site in the endogenous cellular gene with a second zinc finger protein.

65. The method of claim 64, wherein the first and second target sites are adjacent.

66. The method of claim 65, wherein the first and second zinc finger proteins are covalently linked.

67. The method of claim 63, wherein the first zinc finger protein is a fusion protein comprising a regulatory domain.

68. The method of claim 67, wherein the first zinc finger protein is a fusion protein comprising at least two regulatory domains.

69. The method of claim 64, wherein the first and second zinc finger proteins are fusion proteins, each comprising a regulatory domain.

70. The method of claim 69, wherein the first and second zinc finger protein are fusion proteins, each comprising at least two regulatory domains.

71. The method of claim 63, wherein the cell is selected from the group consisting of animal cell, a plant cell, a bacterial cell, a protozoal cell, or a fungal cell.

promoter, and wherein the method further comprises the step of first administering the expression vector to the cell.

52. The method of claim 51, wherein the expression vector is a viral expression vector.

53. The method of claim 51, wherein the expression vector is a retroviral expression vector, an adenoviral vector, or an AAV expression vector.

54. The method of claim 51, wherein the zinc finger protein is encoded by a nucleic acid operably linked to an inducible promoter.

55. The method of claim 51, wherein the zinc finger protein is encoded by a nucleic acid operably linked to a weak promoter.

56. The method of claim 31, wherein the cell comprises less than about $1.5x10^6$ copies of the zinc finger protein.

57. The method of claim 31, wherein the target site is upstream of a transcription initiation site of the endogenous cellular gene.

58. The method of claim 31, wherein the target site is adjacent to a transcription initiation site of the endogenous cellular gene.

59. The method of claim 31, wherein the target site is adjacent to an RNA polymerase pause site downstream of a transcription initiation site of the endogenous cellular gene.

60. The method of claim 31, wherein the zinc finger protein comprises an SP-1 backbone.

61. The method of claim 60, wherein the zinc finger protein comprises a regulatory domain and is humanized.

62. A method of activating expression of an endogenous cellular gene, the method comprising the step of:

41. The method of claim 31, wherein the cell is selected from the group consisting of an animal cell, a plant cell, a bacterial cell, a protozoal cell, or a fungal cell.

42. The method of claim 41, wherein the cell is a mammalian cell.

43. The method of claim 42, wherein the cell is a human cell

44. The method of claim 31, wherein expression of the endogenous cellular gene is activated to at least about 200-500%.

45. The method of claim 31, wherein the endogenous cellular gene is a selected from the group consisting of FAD2-1, EPO, GM-CSF, GDNF, VEGF, and LDL-R.

46. The method of claim 31, wherein the endogenous cellular gene is VEGF.

47. The method of claim 31, wherein the activation of gene expression prevents repression of gene expression.

48. The method of claim 37 or 39, wherein the regulatory domain is selected from the group consisting of a transcriptional activator, or a histone acetyltransferase.

49. The method of claim 31, wherein the method further comprises the step of first administering to the cell a delivery vehicle comprising the zinc finger protein, wherein the delivery vehicle comprises a liposome or a membrane translocation polypeptide.

50. The method of claim 31, wherein the zinc finger protein is encoded by a zinc finger protein nucleic acid operably linked to a promoter, and wherein the method further comprises the step of first administering the nucleic acid to the cell in a lipid:nucleic acid complex or as naked nucleic acid.

51. The method of claim 31, wherein the zinc finger protein is encoded by an expression vector comprising a zinc finger protein nucleic acid operably linked to a

31. A method of activating expression of an endogenous cellular gene, the method comprising the step of:

contacting a first target site in the endogenous cellular gene with a first zinc finger protein, wherein the K_d of the zinc finger protein is less than about 25 nM;

thereby activating expression of the endogenous cellular gene to at least about 150%.

32. The method of claim 31, wherein the endogenous cellular gene is developmentally silent or inactive.

33. The method of claim 32, wherein the endogenous cellular gene is EPO, GATA, hemoglobin gamma, hemoglobin delta, an interleukin, GM-CSF, eutrophin, or MyoD.

34. The method of claim 31, wherein the step of contacting further comprises contacting a second target site in the endogenous cellular gene with a second zinc finger protein.

35. The method of claim 34, wherein the first and second target sites are adjacent.

36. The method of claim 35, wherein the first and second zinc finger proteins are covalently linked.

37. The method of claim 31, wherein the first zinc finger protein is a fusion protein comprising a regulatory domain.

38. The method of claim 37, wherein the first zinc finger protein is a fusion protein comprising at least two regulatory domains.

39. The method of claim 34, wherein the first and second zinc finger proteins are fusion proteins, each comprising a regulatory domain.

40. The method of claim 39, wherein the first and the second zinc finger protein are fusion proteins, each comprising at least two regulatory domains.

21. The method of claim 19, wherein the expression vector is a retroviral expression vector, an adenoviral expression vector, or an AAV expression vector.

22. The method of claim 19, wherein the zinc finger protein is encoded by a nucleic acid operably linked to an inducible promoter.

23. The method of claim 19, wherein the zinc finger protein is encoded by a nucleic acid operably linked to a weak promoter.

24. The method of claim 1, wherein the cell comprises less than about $1.5x10^6$ copies of the zinc finger protein.

25. The method of claim 1, wherein the target site is upstream of a transcription initiation site of the endogenous cellular gene.

26. The method of claim 1, wherein the target site is adjacent to a transcription initiation site of the endogenous cellular gene.

27. The method of claim 1, wherein the target site is adjacent to an RNA polymerase pause site downstream of a transcription initiation site of the endogenous cellular gene.

28. The method of claim 1, wherein the zinc finger protein comprises an SP-1 backbone.

29. The method of claim 28, wherein the zinc finger protein comprises a regulatory domain and is humanized.

30. A method of inhibiting expression of an endogenous cellular gene in a cell, the method comprising the step of:

contacting a target site in the endogenous cellular gene with a fusion zinc finger protein comprising six fingers and a regulatory domain, wherein the K_d of the zinc finger protein is less than about 25 nM;

thereby inhibiting expression of the endogenous cellular gene by at least about 20%.

12. The method of claim 1, wherein expression of the endogenous cellular gene is inhibited by at least about 75%-100%.

13. The method of claim 1, wherein the endogenous cellular gene is a selected from the group consisting of VEGF, ERα, IGF-I, c-myc, c-myb, ICAM, and Her2/Neu.

14. The method of claim 1, wherein the endogenous cellular gene is VEGF.

15. The method of claim 1, wherein the inhibition of gene expression prevents gene activation.

16. The method of claim 5 or 7, wherein the regulatory domain is selected from the group consisting of a transcriptional repressor, an endonuclease, a methyl transferase, and a histone deacetylase.

17. The method of claim 1, wherein the method further comprises the step of first administering to the cell a delivery vehicle comprising the zinc finger protein, wherein the delivery vehicle comprises a liposome or a membrane translocation polypeptide.

18. The method of claim 1, wherein the zinc finger protein is encoded by a zinc finger protein nucleic acid operably linked to a promoter, and wherein the method further comprises the step of first administering the nucleic acid to the cell in a lipid:nucleic acid complex or as naked nucleic acid.

19. The method of claim 1, wherein the zinc finger protein is encoded by an expression vector comprising a zinc finger protein nucleic acid operably linked to a promoter, and wherein the method further comprises the step of first administering the expression vector to the cell.

20. The method of claim 19, wherein the expression vector is a viral expression vector.

WHAT IS CLAIMED IS:

1. A method of inhibiting expression of an endogenous cellular gene in a cell, the method comprising the step of:

contacting a first target site in the endogenous cellular gene with a first zinc finger protein, wherein the K_d of the zinc finger protein is less than about 25 nM;

thereby inhibiting expression of the endogenous cellular gene by at least about 20%.

2. The method of claim 1, wherein the step of contacting further comprises contacting a second target site in the endogenous cellular gene with a second zinc finger protein.

3. The method of claim 2, wherein the first and second target sites are adjacent.

4. The method of claim 3, wherein the first and second zinc finger proteins are covalently linked.

5. The method of claim 1, wherein the first zinc finger protein is a fusion protein comprising a regulatory domain.

6. The method of claim 5, wherein the first zinc finger protein is a fusion protein comprising at least two regulatory domains.

7. The method of claim 2, wherein the first and second zinc finger proteins are fusion proteins, each comprising a regulatory domain.

8. The method of claim 7, wherein the first and second zinc finger protein are fusion proteins, each comprising at least two regulatory domains.

9. The method of claim 1, wherein the cell is selected from the group consisting of animal cell, a plant cell, a bacterial cell, a protozoal cell, or a fungal cell.

10. The method of claim 9, wherein the cell is a mammalian cell

11. The method of claim 10, wherein the cell is a human cell

transactivation domain, as described herein and in USSN 09/229,007, filed January 12, 1999. All plasmid DNAs were prepared using the Qiagen Midi preparation system.

$2x10^6$ Hep3B cells (a human hepatocellular carcinoma-derived cell line) or $5x10^6$ HEK293 cells (a human embryonic kidney epithelium-derived cell line) were seeded into 6-well plates one day before transfection. 500 ng of the effector plasmid (encoding the engineered ZFP) was transiently transfected into the cells using Lipofectamin (GIBCO-BRL). Mock transfection and transfection with an empty expression vector served as controls. One day later the growth medium was removed, and fresh DMEM was added. Culture supernatants were collected 24 hours later for determination of EPO protein expression levels using a commercially available ELISA kit (R&D Systems).

The results in Figure 11a show that transfection of a vector encoding the EPO2C ZFP transactivation protein significantly increased the level of EPO expression when compared to control vector (pcDNANVF) or mock transfected cells. This activation was observed in both Hep3B and HEK293 cells (Figure 11a).

Hep3B cells are known to be capable of expressing the EPO gene under certain conditions (Goldberg *et al., PNAS* 84:7972 (1987)). These cells are derived from hepatocytes, the fetal source of EPO. In Hep3B cells, hypoxia (or treatment with $CoCl_2$, which mimics hypoxia) activates expression of the EPO gene, which is otherwise silent (*see* Figure 11a). VEGF expression is also controlled by hypoxia in Hep3B cells.

Hypoxia or $CoCl_2$ treatment activates VEGF expression in HEK293 cells, showing that the hypoxic response functions normally in this cell line (*see* Figure 8). However, hypoxia or $CoCl_2$ treatment is not able to induce EPO expression in HEK293 cells (Figure 11a). This result occurs because HEK293 cells are derived from embryonic kidney epithelial cells, a tissue that does not act as a source of EPO. Despite inactivation of EPO in these cells, the chimeric EPO2C-VP16 ZFP is able to activate the EPO gene (Figure 11a).

The $CoCl_2$ effects and ZFP effects on EPO gene expression were shown to be effects on EPO transcription. Quantitative RT-PCR was performed using a method known as a TaqMan assay (PE Applied Biosystems). Data are shown in Figure 11b.

10, panel A; *see* Example VII for experimental details). The size of the protected probe was identical to the size of the probe generated from the control human RNA provided as a control for RNA integrity. (Figure 10, panel A).

In a separate experiment, the level of VEGF specific mRNA was also quantitated in cells that had been transfected with a VEGF-KRAB effector plasmid (Figure 10, panel B; *see* Example VI for experimental details). The details of the transfection are described in Example VI. A dramatic decrease in the level of VEGF mRNA was observed when cells were transfected with the VEGF3a/1-KRAB effector plasmid. No significant decrease in VEGF mRNA was observed when cells were transfected with NKF-control or a non-VEGF specific ZFP (CCR5-5-KRAB and CCR5-3-KRAB, which recognize different CCR5 target sites).

This experiment demonstrates that the increase in VEGF protein observed upon transfection with the VEGF1-VP16 chimeric transcription factor is mediated by an increase in the level of VEGF mRNA. Similarly, the decrease in VEGF protein observed upon transfection with the VEGF3a/1-KRAB chimeric transcription factor is mediated by a decrease in the level of VEGF mRNA.

Example IX: Activation of EPO, a developmentally silent gene

EPO is a 30.4 kD glycoprotein hormone and plays a key role in the control of red blood cell production. The EPO gene is normally only expressed in fetal liver and the adult kidney in response to hypoxia. It acts through interacting with the EPO receptor on the erythroid progenitor cells in the bone marrow to stimulate their proliferation and differentiation into mature erythrocytes (Jelkmann, *Physiological Reviews* 72:449(1992); Semenza, *Hematology/Oncology Clinics of North America* 8:863(1994); Ratcliffe *et al.*, *J. Exp. Bio.*(1991)). Recombinant human EPO gene products have been successfully used to treat anemia and chronic renal failure (Egrie, *Pharacotherapy* 10:3S(1990)).

The EPO2C ZFP was designed to recognize a 9-bp DNA-binding site located 853-bp upstream of the EPO transcription initiation site. Methods for design and construction of EPO2C are described herein and in USSN 09/229,007, filed January 12, 1999. The EPO2C binding site sequence is GCGGTGGCT.

Eukaryotic expression vectors were constructed by fusing the sequence encoding EPO2C ZFP to the SV40 nuclear localization signal (NLS) and the HSV VP16

DNA-Lipofectamine mixture was removed, and fresh culture medium containing 10% fetal bovine serum was layered on the cells. One day later, fresh media was added and the supernatant was collected 24 hours later for determination of VEGF levels using a commercially available ELISA kit (R and D Systems).

For the three-fingered VEGF1-specific ZFP (VEGF1-VP16), a 7-10 fold increase in VEGF expression was observed when compared to control plasmid (NVF-control) and mock transfected cells (Figure 9). Similar results have been obtained in 5 independent experiments. It is important to note that the level of VEGF secretion in VEGF1-VP16 transfected cells was equivalent or greater than the level in cells that have been treated with DFX (Figure 9). Introduction of VEGF3a/1-VP16 stimulated a more modest induction of VEGF. This result is consistent with the finding in Example VI, in which expression of the 18-bp binding protein without a functional domain prevented activation to a certain degree. This result suggested that the tight binding of this protein to the start site of transcription interferes with activation.

These data indicate that a designed ZFP is capable of locating and binding to its target site on the chromosome, presenting a transcriptional activation domain, and dramatically enhancing the expression level of that gene. In particular, the results indicate that ZFPs with a K_d of less than about 25 nM (e.g., VEGF1 has an average apparent K_d of about 10 nM) provide dramatic increases in expression.

Example VIII: RNase protection assay

To further substantiate the results in Examples VI and VII, a ribonuclease protection assay (RPA) was performed to correlate the increased level of VEGF protein with an increase in VEGF mRNA levels (Example VII) , and to correlate the decreased level of VEGF protein with a decrease in VEGF mRNA levels (Example VI).

RNA was isolated from the transfected cells using an RNA isolation kit (Pharmingen). Radiolabeled multi template probes, which included a VEGF specific probe, were prepared by *in vitro* transcription and hybridized overnight at 56^0C to 5 µg of each of the RNAs from the experimental and control transfected cells. The hybridization mixture was treated with RNase and the protected probes were purified and subjected to 5% denaturing polyacrylamide gel electrophoresis and the radioactivity was evaluated by autoradiography. 293 cells transfected with the VEGF1-VP16 had a 2-4 fold increase in the level of VEGF mRNA when compared to cells transfected with NVF-control (Figure

expression was observed with VEGF3a/1-NF, which expresses the 18-bp binding protein without a functional domain. This result suggests that binding to the start site of transcription, even without a repression domain, interferes with transcription. Even when fused to the KRAB domain, the VEGF3a ZFP is unable to affect expression levels (plasmid VEGF3a-KRAB). However, VEGF1 fused to KRAB (VEGF1-KRAB) results in a dramatic decrease in expression. VEGF3a/1 fused to KRAB (VEGF3a/1-KRAB) prevents expression of VEGF altogether.

These data indicate that a designed ZFP is capable of locating and binding to its target site on the chromosome and preventing expression of an endogenous cellular target gene. In particular, the results indicate that ZFPs with a K_d of less than about 25 nM (e.g., VEGF1 has an average apparent K_d of about 10 nM) provide dramatic decreases in expression. In addition, the data demonstrate that the KRAB functional domain enhances gene silencing. Because in this experiment the introduction of the repressor occurs before the inducer of VEGF is added (DFX), the data demonstrate the ability of a designed repressor to prevent activation of an already quiescent gene. In addition, these results demonstrate that a six-finger engineered ZFP (VEGF3a/1) with nanomolar affinity for its target is able to inhibit the hypoxic response of the VEGF gene when it binds a target that overlaps the transcriptional start site.

Example VII: Activation of endogenous VEGF gene in human cells

This Example demonstrates that a designed ZFP can activate the expression of a gene that is in its natural context and chromatin structure. Specifically, effector plasmids expressing VEGF ZFPs fused to the VP16 activation domain were introduced into cells and were shown to up-regulate the VEGF gene.

Eucaryotic expression vectors were constructed that fuse the VEGF3a/1 and the VEGF1 ZFPs to the SV40 NLS and VP16, as described in Example III. Transfections were done using Lipofectamine, a commercially available liposome preparation from GIBCO-BRL. All plasmid DNAs were prepared using the Qiagen Midi DNA purification system. 10 μg of the effector plasmid (containing the engineered ZFP) was mixed with 100 μg of Lipofectamine (50 μl) in a total volume of 1600μl of Opti-MEM. A pCMVβ-gal plasmid (Promega) was also included in the DNA mixture as an internal control for transfection efficiency. Following a 30 minute incubation, 6.4 ml of DMEM was added and the mixture was layered on 3 x10^6 293 cells. After five hours, the

efficiency. Following a 30 minute incubation, 6.4 ml of DMEM was added and the mixture was layered on 3 x10^6 293 cells. After five hours, the DNA-Lipofectamine mixture was removed, and fresh culture medium containing 10% fetal bovine serum was layered on the cells.

Eighteen hours post transfection, the 293 cells were induced by treatment with 100 μM DFX (desferrioxamine), resulting in a rapid and lasting transcriptional activation of the VEGF gene and also in a gradual increase in VEGF mRNA stability (Ikeda *et al., J. Biol. Chem.* 270:19761-19766 (1995)). Under routine culture conditions, 293 cells secrete a low level of VEGF in the culture media. The cells were allowed to incubate an additional 24 hours before the supernatants were collected for determination of VEGF levels by an ELISA assay.

In parallel experiments that demonstrated a similar level of repression, cell viability was monitored using the Promega Celltiter 96® Aqueous One Solution cell proliferation assay (Promega). After Dfx treatment for 18 hours, 500 μL of the original 2 ml of media was removed and analyzed for VEGF expression, as described above. To evaluation cell viability, 300 μL of Promega Celltiter 96® Aqueous One Solution Reagent was added to the remaining 1.5 ml. The cells were then incubated at 37°C for approximately 2 hours. 100 μL from each well was transferred to a 96-well plate and read on an ELISA plate reader at OD 490 nm. There was no significant reduction in viability of cells expressing the VEGF3a/1-KRAB construct relative to those transfected with empty vector controls, indicating that the VEGF repression observed was not due to generalized cell death.

A 40-50-fold decrease in VEGF expression was noted in the DFX treated cells transfected with VEGF3a/1-KRAB, an expression vector encoding the 18 bp binding VEGF high affinity ZFP. A two-fold decrease in expression was observed when cells were transfected with VEGF1-KRAB, an expression vector encoding the 9 bp binding VEGF high affinity ZFP. No significant decrease in VEGF expression was observed in cells that were transfected with a non-VEGF ZFP (CCR5-KRAB) or NKF-control (Figure 8). Similar results have been obtained in three independent transfection experiments.

In a separate experiment, the following results were obtained (data not shown). VEGF1-NF, which expresses the 9-bp-binding VEGF1 ZFP without a functional domain, showed no effect on VEGF gene expression. A significant reduction in VEGF

activation relative to the empty pcDNA vector control. VEGF3a/1 (the 18-bp-binding ZFP) expression plasmid activates luciferase expression very substantially, showing about a 14-fold increase relative to pcDNA. These experiments clearly demonstrate that a designed ZFP, when fused to the VP16 activation domain, is capable of functioning in a cell to activate transcription of a gene when its target site is present. Furthermore, these results clearly demonstrate that an 18-bp binding protein, VEGF3a/1, is a much better activator in this assay than a 9-bp binding VEGF1 protein. This could be a result of the improved affinity or decreased off-rate of the VEGF3a/1 protein.

A fourth VEGF reporter plasmid was constructed by cloning the KpnI/NcoI fragment of pVFR2-4x into pGL3-Promoter to create plasmid pVFR4-4x. Activation was observed in co-transfections using this reporter in combination with effector plasmids expressing the VEGF1-VP16 and VEGF3a/1-VP16 fusions (data not shown). This indicates that these artificial trans-activators are functional when bound either upstream or downstream of the start of transcription.

These co-transfection data demonstrate that ZFPs can be used to regulate expression of reporter genes. Such experiments serve as a useful tool for identifying ZFPs for further use as modulators of expression of endogenous cellular genes. As is shown below, modulation results can vary between co-transfection experiments and endogenous gene experiments, while using the same ZFP construct.

Example VI: Repression of an endogenous VEGF gene in human cells

This Example demonstrates that a designed ZFP can repress expression of an endogenous cellular gene that is in its natural context and chromatin structure. Specifically, effector plasmids expressing VEGF ZFPs fused to the KRAB repression domain were introduced into cells and were shown to down-regulate the VEGF gene.

Eucaryotic expression vectors were constructed that fuse the VEGF3a/1 and the VEGF1 ZFPs to the SV40 NLS and KRAB, as described above in Example III. Transfections were done using Lipofectamine, a commercially available liposome preparation from GIBCO-BRL. All plasmid DNAs were prepared using Qiagen Midi DNA purification system. 10 μg of the effector plasmid was mixed with 100 μg of Lipofectamine (50 μl) in a total volume of 1600 μl of Opti-MEM. A pCMVβ-gal plasmid (Promega) was also included in the DNA mixture as an internal control for transfection

above, at the MluI/BglII sites of plasmid pGL3-Promoter (Promega). This vector has been deleted for the SV40 enhancer sequence and therefore has a lower basal level of firefly luciferase expression. pVFR3-4x was constructed by swapping the KpnI/NcoI fragment of pVFR1-4x into the KpnI/NcoI sites of pGL3-Promoter.

The effector plasmid construction is described above. The VEGF1-VP16, VEGF3a-VP16, and VEGF3a/1-VP16 expression vectors were designed to produce a fusion of the SV40 nuclear localization sequence, the VEGF ZFP, the VP16 trans-activation domain, and a FLAG epitope tag all under the control of the CMV promoter. The empty pcDNA3 expression vector was used as a control.

All vectors were prepared using Qiagen DNA purification kits. Figure 7 shows a typical set of transfections using 293 (human embryonic kidney) cells. Approximately 40,000 cells were seeded into each well of a 24-well plate and allowed to grow overnight in D-MEM medium containing 10% fetal bovine serum at 37°C with 5% CO_2. Cells were washed with serum-free D-MEM and overlayed with 200 µl of the same. Plasmids were introduced using a calcium phosphate transfection kit (Gibco-BRL) according to the manufacturer's instructions. Cells in each well were transfected with 1.5 µg of reporter plasmid, 1.5 µg of effector plasmid, and 0.5 µg of an actin/β-gal plasmid. Plasmids were combined with 15 µl of $CaCl_2$ and brought to 100 µl with dH_2O. 100 µl of HEPES solution was added dropwise while vortexing. The mix was incubated for 30 min at room temperature. The 200 µl of calcium phosphate-treated DNA was then added to the medium in each well. Transfections were done in triplicate. After 5 hours, the medium was removed and 1 ml of medium containing 10% serum was added. Cells were harvested 40-48 hours after transfection. Luciferase assays were done using the Dual-Light™ system (Tropix). The third plasmid transfected, actin/β-gal, carries the β-galactosidase gene under the control of the actin promoter and was co-transfected as a standard for transfection efficiency. The β-galactosidase assays were also done according to the manufacturer's protocol (Tropix). The data shown in Figure 7 are the average of triplicate assays normalized against the β-galactosidase activity.

For the control reporter plasmid, pGL3-Promoter (pGL3-P), the presence or absence of the ZFP-VP16 expression plasmid does not significantly influence the luciferase expression level. For pVFR3-4x, the reporter containing four copies of the VEGF target site, presence of VEGF1 (the 9-bp-binding ZFP) shows a very slight

assays were done using the Dual Luciferase™ System (Promega). The third plasmid transfected, pRL-SV40, carries the *Renilla* luciferase gene and was co-transfected as a standard for transfection efficiency. The data shown in Figure 6 are the averages of triplicate assays normalized against the *Renilla* activity.

For the control reporter plasmid pGL3-Control (pGL3-C), the presence or absence of the ZFP-KRAB expression plasmid does not influence the luciferase expression level. However, for pVFR1-4x, the reporter containing four copies of the VEGF target site, presence of the VEGF1 (9-bp-binding ZFP) or VEGF3a/1 (18-bp-binding ZFP) expression plasmid reduces luciferase expression by a factor of 2-3 relative to the empty pcDNA vector control. The VEGF3a (9-bp-binding ZFP) expression plasmid appears to exhibit little or no effect. These experiments clearly demonstrate that a designed ZFP is capable of functioning in a cell to repress transcription of a gene when its target site is present. Furthermore, it appears that a certain level of affinity is required for function; i.e., VEGF1 and VEGF3a/1, with K_ds of 10 nM or less, are functional, whereas VEGF3a, with a K_d of 200 nM, is not.

A second reporter plasmid, pVFR2-4x, was constructed by removing the four copies of the VEGF target sites using HindIII and inserted them into the HindIII site of pGL3-Control (in the forward orientation). This places the target sites between the start site of transcription for the SV40 promoter and the translational start codon of the luciferase gene. In similar co-transfection experiments to those described, approximately 3-4 fold repression of the luciferase signal was observed with the VEGF1-KRAB or VEGF3a/1-KRAB repressors relative to the pcDNA controls (data not shown). This indicates that the repressors are active when bound either upstream or downstream of the start of transcription.

Example V: Activation of VEGF reporters in co-transfection experiments

This Example demonstrates the use of transient co-transfection studies to measure the activity of the ZFP transcriptional activators in cells. The experimental setup is similar to that of Example IV except that a different transfection method, a different cell line, and a different set of reporter and effector plasmids was used.

For activation experiments, a reporter was constructed labeled pVFR3-4x. This reporter contains the four copies of the VEGF targets, with the sequence shown

The reporter plasmid system was based on the pGL3 firefly luciferase vectors (Promega). Four copies of the VEGF target sites were inserted upstream of the SV40 promoter, which is driving the firefly luciferase gene, in the plasmid pGL3-Control to create pVFR1-4x. This plasmid contains the SV40 enhancer and expresses firefly luciferase to high levels in many cell types. Insertions were made by ligating together tandem copies of the two complementary 42-bp oligonucleotides, JVF9 and JVF10, described in Example II. Adaptor sequences were ligated on, and the assembly was inserted into the MluI/BglII sites of pGL3-Control. This resulted in the insertion of the following sequence between those sites:

ACGCGTaagcttGCTAGCGAGCGGGGAGGATCGCGGAGGCTTGGGGCAGCCGGG
TAGAGCGAGCGGGGAGGATCGCGGAGGCTTGGGGCAGCCGGGTAGAGCGAG
CGGGGAGGATCGCGGAGGCTTGGGGCAGCCGGGTAGAGCGAGCGGGGAGGA
TCGCGGAGGCTTGGGGCAGCCGGGTAGAGCGCTCAGaagcttAGATCT.

The first six and last six nucleotides shown are the MluI and BglII sites; the lowercase letters indicate HindIII sites. The binding sites for VEGF1 and VEGF3a are underlined.

The effector plasmid construction is described above. The VEGF1-KRAB, VEGF3a-KRAB, and VEGF3a/1-KRAB expression vectors were designed to produce a fusion of the SV40 nuclear localization sequence, the VEGF ZFP, the KRAB repression domain, and a FLAG epitope marker all under the control of the CMV promoter. The empty pcDNA3.1 expression vector was used as a control (pcDNA).

All vectors were prepared using Qiagen DNA purification kits. Figure 6 shows a typical set of transfections using COS-1 (African green monkey kidney) cells. Approximately 40,000 cells were seeded into each well of a 24-well plate and allowed to grow overnight in Dulbecco's Modified Eagle Medium (D-MEM) medium containing 10% fetal bovine serum at 37°C with 5% CO_2. Cells were washed with PBS and overlayed with 200 µl of serum-free D-MEM. Plasmids were introduced using lipofectamine (Gibco-BRL). Each well was transfected with about 0.3 µg of effector plasmid, 0.3 µg of reporter plasmid, and 0.01 µg of plasmid pRL-SV40 (Promega) that had been complexed with 6 µl of lipofectamine and 25 µl of D-MEM for 30 min at 37°C. Transfections were done in triplicate. After 3 hrs, 1 ml of medium containing 10% serum was added to each well. Cells were harvested 40-48 hours after transfection. Luciferase

ATGGGGTACCCGGGGATGGATCCGGCAGCGACTACAAGGACGACGATGACA AGTAAGCTTCTCGAG

into the EcoRI-XhoI sites of pcDNA-NKF, thereby replacing the NLS-KRAB-FLAG sequences with NLS-FLAG only.

VEGF1-NF and VEGF3a/1-NF were constructed by inserting a KpnI-BamHI cassette containing the ZFP sequences into NF-control digested with KpnI and BamHI. CCR5-KRAB was constructed in the same way as the VEGF KRAB vectors, except that the ZFP sequences were designed to be specific for a DNA target site that is unrelated to the VEGF targets.

Finally, control versions of both the KRAB and VP16 expression plasmids were constructed. Plasmid NKF-control was designed to express NLS-KRAB-FLAG without zinc finger protein sequences; plasmid NVF-control was designed to express NLS-VP16-FLAG without ZFP sequences. These plasmids were made by digesting pcDNA-NKF and -NVF, respectively, with BamHI, filling in the ends with Klenow, and religating in order to place the downstream domains into the proper reading frame. These plasmids serve as rigorous controls for cell culture studies.

Mammalian cell expression and nuclear localization of the VEGF engineered ZFPs was demonstrated through immunofluorescence studies. 293 (human embryonic kidney) cells were transfected with the expression plasmid encoding the NLS-VEGF1-KRAB-FLAG chimera. Lipofectamine was used as described below. After 24-48 hours, cells were fixed and exposed to a primary antibody against the FLAG epitope. A secondary antibody labeled with Texas Red was applied, and the cells were counter stained with DAPI. Texas Red staining was observed to consistently co-localize with the DAPI staining, indicating that the ZFP being expressed from this plasmid was nuclear localized.

Example IV: Repression of VEGF reporters in co-transfection experiments

This Example demonstrates the use of transient co-transfection studies to measure the activity of the ZFP repressor proteins in cells. Such experiments involve co-transfection of ZFP-KRAB expression ("effector") plasmids with reporter plasmids carrying the VEGF target sites. Efficacy is assessed by the repression of reporter gene expression in the presence of the effector plasmid relative to empty vector controls.

Thiesen, 1990, *supra*), the FLAG epitope (from Kodak/IBI catalog), and a HindIII site, altogether comprising the sequence

GGTACCCGGGGATCCCGGACACTGGTGACCTTCAAGGATGTATTTGTGGACTT
CACCAGGGAGGAGTGGAAGCTGCTGGACACTGCTCAGCAGATCGTGTACAGA
AATGTGATGCTGGAGAACTATAAGAACCTGGTTTCCTTGGGCAGCGACTACA
AGGACGACGATGACAAGTAAGCTTCTCGAG

where the KpnI, BamHI and XhoI sites are underlined.

The VEGF3a/1-KRAB effector plasmid was generated by inserting a KpnI-BamHI cassette containing the ZFP sequences into pcDNA-NKF digested with KpnI and BamHI. The VEGF1-KRAB and VEGF3a-KRAB effector plasmids were constructed in a similar way except that the ZFP sequences were first cloned into the NLS-KRAB-FLAG sequences in the context of plasmid pLitmus 28 (New England Biolabs) and subsequently moved to the BamHI-XhoI sites of pcDNA3.1(+) as a BglII-XhoI cassette, where the BglII site was placed immediately upstream of the EcoRI site (see Example IV for expression of these vectors).

The effector plasmids used in Example V were constructed as follows. Plasmid pcDNA-NVF was constructed by PCR amplifying the VP16 transactivation domain, as described above, and inserting the product into the BamHI/HindIII sites of pcDNA-NKF, replacing the KRAB sequence. The sequence of the inserted fragment, from BamHI to HindIII, was:

GGATCCGCCCCCCCGACCGATGTCAGCCTGGGGGACGAGCTCCACTTAGACG
GCGAGGACGTGGCGATGGCGCATGCCGACGCGCTAGACGATTTCGATCTGGA
CATGTTGGGGGACGGGGATTCCCCGGGGCCGGGATTTACCCCCCACGACTCC
GCCCCCTACGGCGCTCTGGATATGGCCGACTTCGAGTTTGAGCAGATGTTTAC
CGATGCCCTTGGAATTGACGAGTACGGTGGGGGCAGCGACTACAAGGACGAC
GATGACAAGTAAGCTT.

VEGF1-VP16 and VEGF3a/1-VP16 vectors were constructed by inserting a KpnI-BamHI cassette containing the ZFP sequences into pcDNA-NVF digested with KpnI and BamHI.

The effector plasmids used in Example VI were constructed as follows. Plasmid NF-control was generated by inserting the sequence

GAATTCGCTAGCGCCACCATGGCCCCCAAGAAGAAGAGGAAGGTGGGAATCC

CGCGGATCCGCCCCCCCGACCGATG, and

(2) JVF25

CCGCAAGCTTACTTGTCATCGTCGTCCTTGTAGTCGCTGCCCCCACCGTACTC GTCAATTCC.

The downstream primer, JVF25, was designed to include a downstream FLAG epitope-encoding sequence.

Three expression vectors were constructed for these studies. The general design is summarized in Figure 5. The vectors are derived from pcDNA3.1(+) (Invitrogen), and place the ZFP constructs under the control of the cytomegalovirus (CMV) promoter. The vector carries ampicillin and neomycin markers for selection in bacteria and mammalian cell culture, respectively. A Kozak sequence for proper translation initiation (Kozak, *J. Biol. Chem.* 266:19867-19870 (1991)) was incorporated. To achieve nuclear localization of the products, the nuclear localization sequence (NLS) from the SV40 large T antigen (Pro-Lys-Lys-Lys-Arg-Lys-Val) (Kalderon *et al., Cell* 39:499-509 (1984)) was added. The insertion site for the ZFP-encoding sequence is followed by the functional domain sequence. The three versions of this vector differ in the functional domain; "pcDNA-NKF" carries the KRAB repression domain sequence, "pcDNA-NVF" carries the VP16 activation domain, and "NF-control" carries no functional domain. Following the functional domain is the FLAG epitope sequence (Kodak) to allow specific detection of the ZFPs.

The vectors were constructed as follows. Plasmid pcDNA-ΔHB was constructed by digesting plasmid pcDNA3.1(+) (Invitrogen) with HindIII and BamHI, filling in the sticky ends with Klenow, and religating. This eliminated the HindIII, KpnI, and BamHI sites in the polylinker. The vector pcDNA3.1(+) is described in the Invitrogen catalog. Plasmid pcDNA-NKF was generated by inserting a fragment into the EcoRI/XhoI sites of pcDNA-ΔHB that contained the following: 1) a segment from EcoRI to KpnI containing the Kozak sequence including the initiation codon and the SV40 NLS sequence, altogether comprising the DNA sequence

GAATTCGCTAGCGCCACCATGGCCCCCAAGAAGAAGAGGAAGGTGGGAATCC ATGGGGTAC,

where the EcoRI and KpnI sites are underlined; and 2) a segment from KpnI to XhoI containing a BamHI site, the KRAB-A box from KOX1 (amino acid coordinates 11-53 in

occupancy of the target site is much higher for the 18-bp binding protein than for the 9-bp binding protein.

Example III: Fusing designed ZFP sequences to functional domains in mammalian expression vectors

This Example describes the development of expression vectors for producing ZFPs within mammalian cells, translocating them to the nucleus, and providing functional domains that are localized to the target DNA sequence by the ZFP. The functional domains employed are the Kruppel-Associated Box (KRAB) repression domain and the Herpes Simplex Virus (HSV-1) VP16 activation domain.

Certain DNA-binding proteins contain separable domains that function as transcriptional repressors. Approximately 20% of ZFPs contain a non-DNA-binding domain of about 90 amino acids that functions as a transcriptional repressor (Thiesen, *The New Biologist* 2:363-374 (1990); Margolin *et al., PNAS* 91:4509-4513 (1994); Pengue *et al.*, (1994), *supra*; Witzgall *et al.*, (1994), *supra*). This domain, termed the KRAB domain, is modular and can be joined to other DNA-binding proteins to block expression of genes containing the target DNA sequence (Margolin *et al.*, (1994); Pengue *et al.*, (1994); Witzgall *et al.*, (1994), *supra*). The KRAB domain has no effect by itself; it needs to be tethered to a DNA sequence via a DNA-binding protein to function as a repressor. The KRAB domain has been shown to block transcription initiation and can function at a distance of up to at least 3 kb from the transcription start site. The KRAB domain from the human KOX-1 protein (Thiesen, *The New Biologist* 2:363-37 (1990)) was used for the studies described here. This 64 amino acid domain can be fused to ZFPs and has been shown to confer repression in cell culture (Liu *et al., supra*).

The VP16 protein of HSV-1 has been studied extensively, and it has been shown that the C-terminal 78 amino acids can act as a trans-activation domain when fused to a DNA-binding domain (Hagmann *et al., J. Virology* 71:5952-5962 (1997)). VP16 has also been shown to function at a distance and in an orientation-independent manner. For these studies, amino acids 413 to 490 in the VP16 protein sequence were used. DNA encoding this domain was PCR amplified from plasmid pMSVP16ΔC+119 using primers with the following sequences:

(1) JVF24

CHIQGCGKVYGTTSNLRRHLRWHTGERPFMCTWSYCGKRFTRSSNLQRHKRTH TGEKKFACPECPKRFMRSDHLSRHIKTHQNKKGGS

The 18-bp binding protein VEGF3a/1 was expressed in *E. coli* as an MBP fusion, purified by affinity chromatography, and tested in EMSA experiments as described in Example I. The target oligonucleotides were prepared as described and comprised the following complementary sequences:

(1) JVF9

AGCGAGCGGGGAGGATCGCGGAGGCTTGGGGCAGCCGGGTAG, and

(2) JVF10

CGCTCTACCCGGCTGCCCCAAGCCTCCGCGATCCTCCCCGCT.

For the EMSA studies, 20 µl binding reactions contained 10 fmole (0.5 nM) 5'-^{32}P-labeled double-stranded target DNA, 35 mM Tris HCl (pH 7.8), 100 mM KCl, 1 mM $MgCl_2$, 5 mM dithiothreitol, 10% glycerol, 20 µg/ml poly dI-dC, 200 µg/ml bovine serum albumin, and 25 µM $ZnCl_2$. Protein was added as one fifth volume from a 3-fold dilution series. Binding was allowed to proceed for 60 min at either room temperature or 37°C. Polyacrylamide gel electrophoresis was carried out at room temperature or 37°C using precast 10% or 10-20% Tris-HCl gels (BioRad) and standard Tris-Glycine running buffer. The room temperature assays yielded an apparent K_d for this VEGF3a/1 protein of approximately 1.5 nM. Thus, the 18-bp binding ZFP bound with high affinity to its target site. In a parallel experiment, VEGF1 protein was tested against its target using the oligonucleotides described in Example I, yielding an apparent K_d of approximately 2.5 nM. When binding and electrophoresis were performed at 37°C, the apparent K_d of VEGF3a/1 was approximately 9 nM when tested against the 18-bp target, compared to a K_d of 40 nM for VEGF1 tested against its target. This indicates that the difference in binding affinities is accentuated at the higher temperature.

The apparent K_d is a useful measure of the affinity of a protein for its DNA target. However, for a DNA binding site either *in vitro* or *in vivo*, its occupancy is determined to a large extent by the off-rate of the DNA-binding protein. This parameter can be measured by competition experiments as shown in Figure 4. The conditions for EMSA were as described above; binding and electrophoresis were performed at 37°C. These data indicate that the half-life of the protein-DNA complex is more than ten times longer for VEGF3a/1 than for VEGF1. Thus, under these *in vitro* conditions, the

at the ends (restriction sites underlined). VEGF1 was PCR amplified using the primers SPEamp13 (5'- GGAGCCAAGGCTGTGGTAAAGTTTACGG) and SPEamp11 (5'- GGAGAAGCTTGGATCCTCATTATCCC) to generate StyI and HindIII restriction sites at the ends (restriction sites underlined). Using synthetic oligonucleotides, the following sequence was ligated between the XbaI and StyI sites, where XbaI and StyI are underlined: TCT AGA CAC ATC AAA ACC CAC CAG AAC AAG AAA GAC GGC GGT GGC AGC GGC AAA AAG AAA CAG CAC ATA TGT CAC ATC CAA GG. This introduced the linker sequence DGGGS between the two SP-1 domains. The ligation product was reamplified with primers SPE7 and SPEamp11 and cloned into pUC19 using the EcoRI and HindIII sites. The linked ZFP sequences were then amplified with primers

(1) GB19

GCCATGCCGGTACCCATACCTGGCAAGAAGAAGCAGCAC)

(2) GB10

CAGATCGGATCCACCCTTCTTATTCTGGTGGGT to introduce KpnI and BamHI sites for cloning into the modified pMAL-c2 expression vector as described above.

The nucleotide sequence of the designed, 6-finger ZFP VEGF3a/1 from KpnI to BamHI is:

GGTACCCATACCTGGCAAGAAGAAGCAGCACATCTGCCACATCCAGGGCTGT
GGTAAAGTTTACGGCCAGTCCTCCGACCTGCAGCGTCACCTGCGCTGGCACAC
CGGCGAGAGGCCTTTCATGTGTACCTGGTCCTACTGTGGTAAACGCTTCACAC
GTTCGTCAAACCTACAGAGGCACAAGCGTACACACACAGGTGAGAAGAAATT
TGCTTGCCCGGAGTGTCCGAAGCGCTTCATGCGAAGTGACGAGCTGTCTAGA
CACATCAAAACCCACCAGAACAAGAAAGACGGCGGTGGCAGCGGCAAAAAG
AAACAGCACATATGTCACATCCAAGGCTGTGGTAAAGTTTACGGCACAACCT
CAAATCTGCGTCGTCACCTGCGCTGGCACACCGGCGAGAGGCCTTTCATGTGT
ACCTGGTCCTACTGTGGTAAACGCTTCACCCGTTCGTCAAACCTGCAGCGTCA
CAAGCGTACCCACACCGGTGAGAAGAAATTTGCTTGCCCGGAGTGTCCGAAG
CGCTTCATGCGTAGTGACCACCTGTCCCGTCACATCAAGACCCACCAGAATAA
GAAGGGTGGATCC

The VEGF3a/1 amino acid translation (using single letter code) is:

VPIPGKKKQHICHIQGCGKVYGQSSDLQRHLRWHTGERPFMCTWSYCGKRFTRS
SNLQRHKRTHTGEKKFACPECPKRFMRSDELSRHIKTHQNKKDGGGSGKKKQHI

the MBP-fused and unfused versions of the proteins bound with similar affinities. K_ds were also determined under these conditions for MBP fusions of the wild-type Zif268 and SP-1 ZFPs, which yielded K_ds of 60 and 65 nM, respectively. These results are similar to binding constants reported in the literature for Zif268 of approximately 2-30 nM (*see, e.g.*, Jamieson *et al., Biochemistry* 33:5689-5695 (1994)). The K_ds for the synthetic VEGF ZFPs therefore compare very favorably with those determined for these naturally-occurring DNA-binding proteins.

In summary, this Example demonstrates the generation of two novel DNA-binding proteins directed to specific targets near the transcriptional start of the VEGF gene. These proteins bind with affinities similar to those of naturally-occurring transcription factors binding to their targets.

Example II: Linking ZFPs to bind an 18-bp target in the human VEGF gene

An important consideration in ZFP design is DNA target length. For random DNA, a sequence of n nucleotides would be expected to occur once every 0.5 x 4^n base-pairs. Thus, DNA-binding domains designed to recognize only 9 bp of DNA would find sites every 130,000 bp and could therefore bind to multiple locations in a complex genome (on the order of 20,000 sites in the human genome). 9-bp putative repressor-binding sequences have been chosen for VEGF in the 5' UTR where they might directly interfere with transcription. However, in case zinc finger domains that recognize 9-bp sites lack the necessary affinity or specificity when expressed inside cells, a larger domain was constructed to recognize 18 base-pairs by joining separate three-finger domains with a linker sequence to form a six-finger protein. This should ensure that the repressor specifically targets the appropriate sequence, particularly under conditions where only small amounts of the repressor are being produced. The 9-bp target sites in VEGF were chosen to be adjacent to one another so that the zinc fingers could be linked to recognize an 18-bp sequence. The linker DGGGS was chosen because it permits binding of ZFPs to two 9-bp sites that are separated by a one nucleotide gap, as is the case for the VEGF1 and VEGF3a sites (*see also* Liu *et al., PNAS* 5525-5530 (1997)).

The 6-finger VEGF3a/1 protein encoding sequence was generated as follows. VEGF3a was PCR amplified using the primers SPE7 (5'-GAGCAGAATTCGGCAAGAAGAAGCAGCAC) and SPEamp12 (5'-GTGGTCTAGACAGCTCGTCACTTCGC) to generate EcoRI and XbaI restriction sites

VEGF site 3, top: 5'-CATGCATATCGCGGAGGCTTGGCATCGAT

VEGF site 3, bottom: 5'-ATCGATGCCAAGCCTCCGCGATATGCATG

The VEGF DNA target sites are underlined. The 3 bp on either side of the 9 bp binding site was also derived from the actual VEGF DNA sequence. The top strand of each target site was labeled with polynucleotide kinase and γ-^{32}P dATP. Top and bottom strands were annealed in a reaction containing each oligonucleotide at 0.5 μM, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 50 mM NaCl. The mix was heated to 95°C for 5 min. and slow cooled to 30°C over 60 min. Duplex formation was confirmed by polyacrylamide gel electrophoresis. Free label and ssDNA remaining in the target preparations did not appear to interfere with the binding reactions.

Binding of the ZFPs to target oligonucleotides was performed by titrating protein against a fixed amount of duplex substrate. Twenty microliter binding reactions contained 10 fmole (0.5 nM) 5'-^{32}P-labeled double-stranded target DNA, 35 mM Tris HCl (pH 7.8), 100 mM KCl, 1 mM $MgCl_2$, 1 mM dithiothreitol, 10% glycerol, 20 μg/ml poly dI-dC (optionally), 200 μg/ml bovine serum albumin, and 25 μM $ZnCl_2$. Protein was added as one fifth volume from a dilution series made in 200 mM NaCl, 20 mM Tris (pH 7.5), 1 mM DTT. Binding was allowed to proceed for 30 min. at room temperature. Polyacrylamide gel electrophoresis was carried out at 4°C using precast 10% or 10-20% Tris-HCl gels (BioRad) and standard Tris-Glycine running buffer containing 0.1 mM $ZnCl_2$.

The results of a typical EMSA using an MBP fused ZFP are shown in Figure 3. In this case, a 3-fold dilution series of the MBP-VEGF1 protein was used. The shifted product was quantitated on a phosphorimager (Molecular Dynamics) and the relative signal (percent of plateau value) vs. the $\log_{10}$ of nM protein concentration was plotted. An apparent K_d was found by determining the protein concentration that gave half maximal binding of MBP-VEGF1 to its target site, which in this experiment was approximately 2 nM.

The binding affinities determined for the VEGF proteins can be summarized as follows. VEGF1 showed the stronger DNA-binding affinity; in multiple EMSA analyses, the average apparent K_d was determined to be approximately 10 nM when bound to VEGF site 1. VEGF3a bound well to its target site but with a higher apparent K_d than VEGF1; the average K_d for VEGF3a was about 200 nM. In both cases

pure as estimated by staining of SDS polyacrylamide gels with Coomassie blue, and the product migrated at the predicted molecular weight of around 11 kDa (Figure 2).

The second method of producing ZFPs was to express them as fusions to the *E. coli* Maltose Binding Protein (MBP). N-terminal MBP fusions to the ZFPs were constructed by PCR amplification of the pET15b clones and insertion into the vector pMal-c2 under the control of the Tac promoter (New England Biolabs). The fusion allows simple purification and detection of the recombinant protein. It had been reported previously that zinc finger DNA-binding proteins can be expressed from this vector in soluble form to high levels in *E. coli* and can bind efficiently to the appropriate DNA target without refolding (Liu *et al. PNAS* 94:5525-5530 (1997)). Production of MBP-fused proteins was as described by the manufacturer (New England Biolabs). Transformants were grown in LB medium supplemented with glucose and ampicillin, and were induced with IPTG for 3 hrs at 37°C. The cells were lysed by French press, then exposed to an agarose-based amylose resin, which specifically binds to the MBP moiety, thus acting as an affinity resin for this protein. The MBP fusion protein was eluted with 10 mM maltose (Figure 2C) to release ZFP of >50% purity. In some cases, the proteins were further concentrated using a Centricon 30 filter unit (Amicon).

Partially purified unfused and MBP fusion ZFPs were tested by EMSA to assess binding to their target DNA sequences. The protein concentrations in the preparations were measured by Bradford assay (BioRad). Since SDS polyacrylamide gels demonstrated >50% homogeneity by either purification method, no adjustment was made for ZFP purity in the calculations. In addition, there could be significant amounts of inactive protein in the preparations. Therefore, the data generated by EMSAs below represent an underestimate of the true affinity of the proteins for their targets (i.e., overestimate of K_ds). Two separate preparations were made for each protein to help control for differences in ZFP activity.

The VEGF DNA target sites for the EMSA experiments were generated by embedding the 9-bp binding sites in 29-bp duplex oligonucleotides. The sequences of the recognition ("top") strand and their complements ("bottom") used in the assays are as follows:

VEGF site 1, top: 5'-CATGCATAGCGGGGAGGATCGCCATCGAT

VEGF site 1, bottom: 5'-ATCGATGGCGATCCTCCCCGCTATGCATG

TGCTTGCCCGGAGTGTCCGAAGCGCTTCATGCGTAGTGACCACCTGTCCCGTC
ACATCAAGACCCACCAGAATAAGAAGGGTGGATCC

VEGF1 translation:

VPIPGKKKQHICHIQGCGKVYGTTSNLRRHLRWHTGERPFMCTWSYCGKRFTRS
SNLQRHKRTHTGEKKFACPECPKRFMRSDHLSRHIKTHQNKKGGS

VEGF3a:

GGTACCCATACCTGGCAAGAAGAAGCAGCACATCTGCCACATCCAGGGCTGT
GGTAAAGTTTACGGCCAGTCCTCCGACCTGCAGCGTCACCTGCGCTGGCACAC
CGGCGAGAGGCCTTTCATGTGTACCTGGTCCTACTGTGGTAAACGCTTCACCC
GTTCGTCAAACCTACAGAGGCACAAGCGTACACACACCGGTGAGAAGAAATT
TGCTTGCCCGGAGTGTCCGAAGCGCTTCATGCGAAGTGACGAGCTGTCACGA
CATATCAAGACCCACCAGAACAAGAAGGGTGGATCC

VEGF3a translation:

VPIPGKKKQHICHIQGCGKVYGQSSDLQRHLRWHTGERPFMCTWSYCGKRFTRS
SNLQRHKRTHTGEKKFACPECPKRFMRSDELSRHIKTHQNKKGGS

The ability of the designed ZFPs to bind their target sites was verified by expressing and purifying recombinant protein from *E. coli* and performing electrophoretic mobility shift assays (EMSAs). The expression of ZFPs was carried out in two different systems. In the first, the DNA-binding peptides were expressed in *E. coli* by inserting them into the commercially available pET15b vector (Novagen). This vector contains a T7 promoter sequence to drive expression of the recombinant protein. The constructs were introduced into *E. coli* BL21/DE3 ($lacI^q$) cells, which contain an IPTG-inducible T7 polymerase. Cultures were supplemented with 50 μM $ZnCl_2$, were grown at 37°C to an OD at 600 nm of 0.5-0.6, and protein production was induced with IPTG for 2 hrs. ZFP expression was seen at very high levels, approximately 30% of total cellular protein (Figure 2). These proteins are referred to as "unfused" ZFPs.

Partially pure unfused ZFPs were produced as follows (adapted from Desjarlais & Berg, *Proteins: Structure, Function and Genetics* 12:101-104 (1992)). A frozen cell pellet was resuspended in 1/50th volume of 1 M NaCl, 25 mM Tris HCl (pH 8.0), 100 μM $ZnCl_2$, 5 mM DTT. The samples were boiled for 10 min. and centrifuged for 10 min. at ~3,000 x g. At this point the ZFP protein in the supernatant was > 50%

The selection of amino acids in the recognition helices of the two designed ZFPs, VEGF1 and VEGF3a, is summarized in Table 1.

Table 1

Amino acids chosen for recognition helices of VEGF-recognizing ZFPs

Position:	Finger 1				Finger 2				Finger 3			
Protein	-1	2	3	6	-1	2	3	6	-1	2	3	6
VEGF1	T	S	N	R	R	S	N	R	R	D	H	R
VEGF3A	Q	S	D	R	R	S	N	R	R	D	E	R

Coding sequences were constructed to express these peptides using a PCR-based assembly procedure that utilizes six overlapping oligonucleotides (Figure 1). Three oligonucleotides (oligos 1, 3, and 5 in Figure 1) corresponding to "universal" sequences that encode portions of the DNA-binding domain between the recognition helices. These oligonucleotides remain constant for any zinc finger construct. The other three "specific" oligonucleotides (oligos 2, 4, and 6 in Figure 1) were designed to encode the recognition helices. These oligonucleotides contained substitutions at positions -1, 2, 3 and 6 on the recognition helices to make them specific for each of the different DNA-binding domains. Codon bias was chosen to allow expression in both mammalian cells and *E. coli*.

The PCR synthesis was carried out in two steps. First, the double stranded DNA template was created by combining the six oligonucleotides (three universal, three specific) and using a four cycle PCR reaction with a low temperature (25°) annealing step. At this temperature, the six oligonucleotides join to form a DNA "scaffold." The gaps in the scaffold were filled in by a combination of Taq and Pfu polymerases. In the second phase of construction, the zinc finger template was amplified in thirty cycles by external primers that were designed to incorporate restriction sites for cloning into pUC19. Accuracy of clones for the VEGF ZFPs were verified by DNA sequencing. The DNA sequences of each of the two constructs are listed below.

VEGF1:

GGTACCCATACCTGGCAAGAAGAAGCAGCACATCTGCCACATCCAGGGCTGT
GGTAAAGTTTACGGCACAACCTCAAATCTGCGTCGTCACCTGCGCTGGCACAC
CGGCGAGAGGCCTTTCATGTGTACCTGGTCCTACTGTGGTAAACGCTTCACCC
GTTCGTCAAACCTGCAGCGTCACAAGCGTACCCACACCGGTGAGAAGAAATT

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

EXAMPLES

The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of noncritical parameters that could be changed or modified to yield essentially similar results.

Example I: Design and testing of ZFPs targeted to the human VEGF gene

This first Example demonstrates the construction of ZFPs designed to recognize DNA sequences contained in the promoter of the human vascular endothelial growth factor (VEGF) gene. VEGF is an approximately 46 kDa glycoprotein that is an endothelial cell-specific mitogen induced by hypoxia. VEGF has been implicated in angiogenesis associated with cancer, various retinopathies, and other serious diseases. The DNA target site chosen was a region surrounding the transcription initiation site of the gene. The two 9 base pair (bp) sites chosen are found within the sequence agcGGGGAGGATcGCGGAGGCTtgg, where the upper-case letters represent actual 9-bp targets. The protein targeting the upstream 9-bp target was denoted VEGF1, and the protein targeting the downstream 9-bp target was denoted VEGF3a. The major start site of transcription for VEGF is at the T at the 3' end of the first 9-bp target, which is underlined in the sequence above.

The human SP-1 transcription factor was used as a progenitor molecule for the construction of designed ZFPs. SP-1 has a three finger DNA-binding domain related to the well-studied murine Zif268 (Christy *et al.*, *PNAS* 85:7857-7861 (1988)). Site-directed mutagenesis experiments using this domain have shown that the proposed "recognition rules" that operate in Zif268 can be used to adapt SP-1 to other target DNA sequences (Desjarlais & Berg, *PNAS* 91:11099-11103 (1994)). The SP-1 sequence used for construction of zinc finger clones corresponds to amino acids 533 to 624 in the SP-1 transcription factor.

Selection or induction during the course of development can in some cases trigger the substitution of one family member for another mutant member. This type of functional substitution may not be possible in the adult animal. A typical result of developmental compensation would be the lack of a phenotype in a knockout mouse when the ablation of that gene's function in an adult would otherwise cause a physiological change. This is a kind of false negative result that often confounds the interpretation of conventional knockout mouse models.

A few new methods have been developed to avoid embryonic lethality. These methods are typified by an approach using the cre recombinase and lox DNA recognition elements. The recognition elements are inserted into a gene of interest using homologous recombination (as described above) and the expression of the recombinase induced in adult mice post-development. This causes the deletion of a portion of the target gene and avoids developmental complications. The method is labor intensive and suffers form chimerism due to non-uniform induction of the recombinase.

The use of engineered ZFPs to manipulate gene expression can be restricted to adult animals using the small molecule regulated systems described in the previous section. Expression and/or function of a zinc finger-based repressor can be switched off during development and switched on at will in the adult animals. This approach relies on the addition of the ZFP expressing module only; homologous recombination is not required. Because the ZFP repressors are *trans* dominant, there is no concern about germline transmission or homozygosity. These issues dramatically affect the time and labor required to go from a poorly characterized gene candidate (a cDNA or EST clone) to a mouse model. This ability can be used to rapidly identify and/or validate gene targets for therapeutic intervention, generate novel model systems and permit the analysis of complex physiological phenomena (development, hematopoiesis, transformation, neural function etc.). Chimeric targeted mice can be derived according to Hogan *et al.*, *Manipulating the Mouse Embryo: A Laboratory Manual*, (1988); *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed., (1987); and Capecchi *et al.*, *Science* 244:1288 (1989.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

gene of interest or any transgene by placing the function and/or expression of a ZFP regulator under small molecule control.

Transgenic mice

A further application of the ZFP technology is manipulating gene expression in transgenic animals. As with cell lines, over-expression of an endogenous gene or the introduction of a heterologous gene to a transgenic animal, such as a transgenic mouse, is a fairly straightforward process. The ZFP technology is an improvement in these types of methods because one can circumvent the need for generating full-length cDNA clones of the gene under study.

Likewise, as with cell-based systems, conventional down-regulation of gene expression in transgenic animals is plagued by technical difficulties. Gene knockout by homologous recombination is the method most commonly applied currently. This method requires a relatively long genomic clone of the gene to be knocked out (ca. 10 kb). Typically, a selectable marker is inserted into an exon of the gene of interest to effect the gene disruption, and a second counter-selectable marker provided outside of the region of homology to select homologous versus non-homologous recombinants. This construct is transfected into embryonic stem cells and recombinants selected in culture. Recombinant stem cells are combined with very early stage embryos generating chimeric animals. If the chimerism extends to the germline homozygous knockout animals can be isolated by back-crossing. When the technology is successfully applied, knockout animals can be generated in approximately one year. Unfortunately two common issues often prevent the successful application of the knockout technology; embryonic lethality and developmental compensation. Embryonic lethality results when the gene to be knocked out plays an essential role in development. This can manifest itself as a lack of chimerism, lack of germline transmission or the inability to generate homozygous back crosses. Genes can play significantly different physiological roles during development versus in adult animals. Therefore, embryonic lethality is not considered a rationale for dismissing a gene target as a useful target for therapeutic intervention in adults. Embryonic lethality most often simply means that the gene of interest can not be easily studied in mouse models, using conventional methods.

Developmental compensation is the substitution of a related gene product for the gene product being knocked out. Genes often exist in extensive families.

difficult. Until now, simple methods for assigning function to differentially expressed genes have not kept pace with the ability to monitor differential gene expression.

Using conventional molecular approaches, over expression of a candidate gene can be accomplished by cloning a full-length cDNA, subcloning it into a mammalian expression vector and transfecting the recombinant vector into an appropriate host cell. This approach is straightforward but labor intensive, particularly when the initial candidate gene is represented by a simple expressed sequence tag (EST). Under expression of a candidate gene by "conventional" methods is yet more problematic. Antisense methods and methods that rely on targeted ribozymes are unreliable, succeeding for only a small fraction of the targets selected. Gene knockout by homologous recombination works fairly well in recombinogenic stem cells but very inefficiently in somatically derived cell lines. In either case large clones of syngeneic genomic DNA (on the order of 10 kb) should be isolated for recombination to work efficiently.

The ZFP technology can be used to rapidly analyze differential gene expression studies. Engineered ZFPs can be readily used to up or down-regulate any endogenous target gene. Very little sequence information is required to create a gene-specific DNA binding domain. This makes the ZFP technology ideal for analysis of long lists of poorly characterized differentially expressed genes. One can simply build a zinc finger-based DNA binding domain for each candidate gene, create chimeric up and down-regulating artificial transcription factors and test the consequence of up or down-regulation on the phenotype under study (transformation, response to a cytokine etc.) by switching the candidate genes on or off one at a time in a model system.

This specific example of using engineered ZFPs to add functional information to genomic data is merely illustrative. Any experimental situation that could benefit from the specific up or down-regulation of a gene or genes could benefit from the reliability and ease of use of engineered ZFPs.

Additionally, greater experimental control can be imparted by ZFPs than can be achieved by more conventional methods. This is because the production and/or function of an engineered ZFP can be placed under small molecule control. Examples of this approach are provided by the Tet-On system, the ecdysone-regulated system and a system incorporating a chimeric factor including a mutant progesterone receptor. These systems are all capable of indirectly imparting small molecule control on any endogenous

Agrobacterium tumefaciens-meditated transformation techniques are well described in the scientific literature (*see, e.g.*, Horsch *et al. Science* 233:496-498 (1984)); and Fraley *et al. PNAS* 80:4803 (1983)).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired ZFP-controlled phenotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the ZFP nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans *et al., Protoplasts Isolation and Culture, Handbook of Plant Cell Culture*, pp. 124-176 (1983); and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73 (1985). Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee *et al. Ann. Rev. of Plant Phys.* 38:467-486 (1987).

Functional genomics assays

ZFPs also have use for assays to determine the phenotypic consequences and function of gene expression. The recent advances in analytical techniques, coupled with focussed mass sequencing efforts have created the opportunity to identify and characterize many more molecular targets than were previously available. This new information about genes and their functions will speed along basic biological understanding and present many new targets for therapeutic intervention. In some cases analytical tools have not kept pace with the generation of new data. An example is provided by recent advances in the measurement of global differential gene expression. These methods, typified by gene expression microarrays, differential cDNA cloning frequencies, subtractive hybridization and differential display methods, can very rapidly identify genes that are up or down-regulated in different tissues or in response to specific stimuli. Increasingly, such methods are being used to explore biological processes such as, transformation, tumor progression, the inflammatory response, neurological disorders etc. One can now very easily generate long lists of differentially expressed genes that correlate with a given physiological phenomenon, but demonstrating a causative relationship between an individual differentially expressed gene and the phenomenon is

region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill.

Alternatively, the plant promoter may direct expression of the ZFP in a specific tissue or may be otherwise under more precise environmental or developmental control. Such promoters are referred to here as "inducible" promoters. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions or the presence of light.

Examples of promoters under developmental control include promoters that initiate transcription only in certain tissues, such as fruit, seeds, or flowers. For example, the use of a polygalacturonase promoter can direct expression of the ZFP in the fruit, a CHS-A (chalcone synthase A from petunia) promoter can direct expression of the ZFP in flower of a plant.

The vector comprising the ZFP sequences will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosluforon or Basta.

Such DNA constructs may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski *et al. EMBO J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm *et al. PNAS* 82:5824 (1985). Ballistic transformation techniques are described in Klein *et al. Nature* 327:70-73 (1987).

number of double bonds, or degree of saturation, determines the melting temperature, reactivity, cooking performance, and health attributes of the resulting oil.

The enzyme responsible for the conversion of oleic acid (18:1) into linoleic acid (18:2) (which is then the precursor for 18:3 formation) is Δ12-oleate desaturase, also referred to as omega-6 desaturase. A block at this step in the fatty acid desaturation pathway should result in the accumulation of oleic acid at the expense of polyunsaturates.

In one embodiment ZFPs are used to regulate expression of the FAD2-1 gene in soybeans. Two genes encoding microsomal Δ6 desaturases have been cloned recently from soybean, and are referred to as FAD2-1 and FAD2-2 (Heppard *et al., Plant Physiol.* 110:311-319 (1996)). FAD2-1 (delta 12 desaturase) appears to control the bulk of oleic acid desaturation in the soybean seed. ZFPs can thus be used to modulate gene expression of FAD2-1 in plants. Specifically, ZFPs can be used to inhibit expression of the FAD2-1 gene in soybean in order to increase the accumulation of oleic acid (18:1) in the oil seed. Moreover, ZFPs can be used to modulate expression of any other plant gene, such as delta-9 desaturase, delta-12 desaturases from other plants, delta-15 desaturase, acetyl-CoA carboxylase, acyl-ACP-thioesterase, ADP-glucose pyrophosphorylase, starch synthase, cellulose synthase, sucrose synthase, senescence-associated genes, heavy metal chelators, fatty acid hydroperoxide lyase, polygalacturonase, EPSP synthase, plant viral genes, plant fungal pathogen genes, and plant bacterial pathogen genes.

Recombinant DNA vectors suitable for transformation of plant cells are also used to deliver the ZFP of the invention to plant cells. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature (*see, e.g.*, Weising *et al. Ann. Rev. Genet.* 22:421-477 (1988)). A DNA sequence coding for the desired ZFP is combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the ZFP in the intended tissues of the transformed plant.

For example, a plant promoter fragment may be employed which will direct expression of the ZFP in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation

non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials. Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described.

Regulation of gene expression in plants

ZFPs can be used to engineer plants for traits such as increased disease resistance, modification of structural and storage polysaccharides, flavors, proteins, and fatty acids, fruit ripening, yield, color, nutritional characteristics, improved storage capability, and the like. In particular, the engineering of crop species for enhanced oil production, e.g., the modification of the fatty acids produced in oilseeds, is of interest.

Seed oils are composed primarily of triacylglycerols (TAGs), which are glycerol esters of fatty acids. Commercial production of these vegetable oils is accounted for primarily by six major oil crops (soybean, oil palm, rapeseed, sunflower, cotton seed, and peanut.) Vegetable oils are used predominantly (90%) for human consumption as margarine, shortening, salad oils, and frying oil. The remaining 10% is used for non-food applications such as lubricants, oleochemicals, biofuels, detergents, and other industrial applications.

The desired characteristics of the oil used in each of these applications varies widely, particularly in terms of the chain length and number of double bonds present in the fatty acids making up the TAGs. These properties are manipulated by the plant in order to control membrane fluidity and temperature sensitivity. The same properties can be controlled using ZFPs to produce oils with improved characteristics for food and industrial uses.

The primary fatty acids in the TAGs of oilseed crops are 16 to 18 carbons in length and contain 0 to 3 double bonds. Palmitic acid (16:0 [16 carbons: 0 double bonds]), oleic acid (18:1), linoleic acid (18:2), and linolenic acid (18:3) predominate. The

rheumatoid arthritis, psoriasis, HIV infection, sickle cell anemia, Alzheimer's disease, muscular dystrophy, neurodegenerative diseases, vascular disease, cystic fibrosis, stroke, and the like. Examples of microorganisms that can be inhibited by ZFP gene therapy include pathogenic bacteria, e.g., chlamydia, rickettsial bacteria, mycobacteria, staphylococci, streptococci, pneumococci, meningococci and conococci, klebsiella, proteus, serratia, pseudomonas, legionella, diphtheria, salmonella, bacilli, cholera, tetanus, botulism, anthrax, plague, leptospirosis, and Lyme disease bacteria; infectious fungus, e.g., *Aspergillus*, *Candida* species; protozoa such as sporozoa (e.g., *Plasmodia*), rhizopods (e.g., *Entamoeba*) and flagellates (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.);viral diseases, e.g., hepatitis (A, B, or C), herpes virus (e.g., VZV, HSV-1, HSV-6, HSV-II, CMV, and EBV), HIV, Ebola, adenovirus, influenza virus, flaviviruses, echovirus, rhinovirus, coxsackie virus, cornovirus, respiratory syncytial virus, mumps virus, rotavirus, measles virus, rubella virus, parvovirus, vaccinia virus, HTLV virus, dengue virus, papillomavirus, poliovirus, rabies virus, and arboviral encephalitis virus, etc.

Administration of therapeutically effective amounts is by any of the routes normally used for introducing ZFP into ultimate contact with the tissue to be treated. The ZFPs are administered in any suitable manner, preferably with pharmaceutically acceptable carriers. Suitable methods of administering such modulators are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention (*see, e.g., Remington's Pharmaceutical Sciences*, 17th ed. 1985)).

The ZFPs, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

Formulations suitable for parenteral administration, such as, for example, by intravenous, intramuscular, intradermal, and subcutaneous routes, include aqueous and

competition for ZFP binding by other sites. This calculation also assumes that essentially all of the ZFP is localized to the nucleus. A value of 100x K_d is used to calculate approximately 99% binding of to the target site, and a value of 10x K_d is used to calculate approximately 90% binding of to the target site. For this example, K_d = 25 nM

ZFP + target site ↔ complex

i.e., DNA + protein ↔ DNA:protein complex

$$K_d = \frac{[\text{DNA}]\,[\text{protein}]}{[\text{DNA:protein complex}]}$$

When 50% of ZFP is bound, K_d = [protein]

So when [protein] = 25 nM and the nucleus volume is 10^{-12} L

[protein] = (25×10^{-9} moles/L) (10^{-12} L/nucleus) (6×10^{23} molecules/mole)

= 15,000 molecules/nucleus for 50% binding

When 99% target is bound; 100x K_d = [protein]

100x K_d = [protein] = 2.5 μM

(2.5×10^{-6} moles/L) (10^{-12}L/nucleus) (6×10^{23} molecules/mole)

= about 1,500,000 molecules per nucleus for 99% binding of target site.

The appropriate dose of an expression vector encoding a ZFP can also be calculated by taking into account the average rate of ZFP expression from the promoter and the average rate of ZFP degradation in the cell. Preferably, a weak promoter such as a wild-type or mutant HSV TK is used, as described above. The dose of ZFP in micrograms is calculated by taking into account the molecular weight of the particular ZFP being employed.

In determining the effective amount of the ZFP to be administered in the treatment or prophylaxis of disease, the physician evaluates circulating plasma levels of the ZFP or nucleic acid encoding the ZFP, potential ZFP toxicities, progression of the disease, and the production of anti-ZFP antibodies. Administration can be accomplished via single or divided doses.

Pharmaceutical compositions and administration

ZFPs and expression vectors encoding ZFPs can be administered directly to the patient for modulation of gene expression and for therapeutic or prophylactic applications, for example, cancer, ischemia, diabetic retinopathy, macular degeneration,

carcinoembryonic antigen (CEA). Sites of viral infection can be diagnosed using various viral antigens such as hepatitis B core and surface antigens (HBVc, HBVs) hepatitis C antigens, Epstein-Barr virus antigens, human immunodeficiency type-1 virus (HIV1) and papilloma virus antigens. Inflammation can be detected using molecules specifically recognized by surface molecules which are expressed at sites of inflammation such as integrins (e.g., VCAM-1), selectin receptors (e.g., ELAM-1) and the like.

Standard methods for coupling targeting agents to liposomes can be used. These methods generally involve incorporation into liposomes lipid components, e.g., phosphatidylethanolamine, which can be activated for attachment of targeting agents, or derivatized lipophilic compounds, such as lipid derivatized bleomycin. Antibody targeted liposomes can be constructed using, for instance, liposomes which incorporate protein A (*see* Renneisen *et al.*, *J. Biol. Chem.*, 265:16337-16342 (1990) and Leonetti *et al.*, *PNAS* 87:2448-2451 (1990).

Doses of ZFPs

For therapeutic applications of ZFPs, the dose administered to a patient, in the context of the present invention should be sufficient to effect a beneficial therapeutic response in the patient over time. In addition, particular dosage regimens can be useful for determining phenotypic changes in an experimental setting, e.g., in functional genomics studies, and in cell or animal models. The dose will be determined by the efficacy and K_d of the particular ZFP employed, the nuclear volume of the target cell, and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular patient.

The maximum therapeutically effective dosage of ZFP for approximately 99% binding to target sites is calculated to be in the range of less than about $1.5x10^5$ to $1.5x10^6$ copies of the specific ZFP molecule per cell. The number of ZFPs per cell for this level of binding is calculated as follows, using the volume of a HeLa cell nucleus (approximately 1000 μm^3 or 10^{-12} L; *Cell Biology*, (Altman & Katz, eds. (1976)). As the HeLa nucleus is relatively large, this dosage number is recalculated as needed using the volume of the target cell nucleus. This calculation also does not take into account

involves using an agent to induce a permeability change in the liposome vesicle. Liposome membranes can be constructed so that they become destabilized when the environment becomes acidic near the liposome membrane (*see, e.g., PNAS* 84:7851 (1987); *Biochemistry* 28:908 (1989)). When liposomes are endocytosed by a target cell, for example, they become destabilized and release their contents. This destabilization is termed fusogenesis. Dioleoylphosphatidylethanolamine (DOPE) is the basis of many "fusogenic" systems.

Such liposomes typically comprise a ZFP and a lipid component, e.g., a neutral and/or cationic lipid, optionally including a receptor-recognition molecule such as an antibody that binds to a predetermined cell surface receptor or ligand (e.g., an antigen). A variety of methods are available for preparing liposomes as described in, e.g., Szoka *et al., Ann. Rev. Biophys. Bioeng.* 9:467 (1980), U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,946,787, PCT Publication No. WO 91\17424, Deamer & Bangham, *Biochim. Biophys. Acta* 443:629-634 (1976); Fraley, *et al., PNAS* 76:3348-3352 (1979); Hope *et al., Biochim. Biophys. Acta* 812:55-65 (1985); Mayer *et al., Biochim. Biophys. Acta* 858:161-168 (1986); Williams *et al., PNAS* 85:242-246 (1988); *Liposomes* (Ostro (ed.), 1983, Chapter 1); Hope *et al., Chem. Phys. Lip.* 40:89 (1986); Gregoriadis, *Liposome Technology* (1984) and Lasic, *Liposomes: from Physics to Applications* (1993)). Suitable methods include, for example, sonication, extrusion, high pressure/homogenization, microfluidization, detergent dialysis, calcium-induced fusion of small liposome vesicles and ether-fusion methods, all of which are well known in the art.

In certain embodiments of the present invention, it is desirable to target the liposomes of the invention using targeting moieties that are specific to a particular cell type, tissue, and the like. Targeting of liposomes using a variety of targeting moieties (e.g., ligands, receptors, and monoclonal antibodies) has been previously described (*see, e.g.*, U.S. Patent Nos. 4,957,773 and 4,603,044).

Examples of targeting moieties include monoclonal antibodies specific to antigens associated with neoplasms, such as prostate cancer specific antigen and MAGE. Tumors can also be diagnosed by detecting gene products resulting from the activation or over-expression of oncogenes, such as ras or c-erbB2. In addition, many tumors express antigens normally expressed by fetal tissue, such as the alphafetoprotein (AFP) and

Toxin molecules also have the ability to transport polypeptides across cell membranes. Often, such molecules are composed of at least two parts (called "binary toxins"): a translocation or binding domain or polypeptide and a separate toxin domain or polypeptide. Typically, the translocation domain or polypeptide binds to a cellular receptor, and then the toxin is transported into the cell. Several bacterial toxins, including *Clostridium perfringens* iota toxin, diphtheria toxin (DT), *Pseudomonas* exotoxin A (PE), pertussis toxin (PT), *Bacillus anthracis* toxin, and pertussis adenylate cyclase (CYA), have been used in attempts to deliver peptides to the cell cytosol as internal or amino-terminal fusions (Arora *et al., J. Biol. Chem.*, 268:3334-3341 (1993); Perelle *et al., Infect. Immun.*, 61:5147-5156 (1993); Stenmark *et al., J. Cell Biol.* 113:1025-1032 (1991); Donnelly *et al., PNAS* 90:3530-3534 (1993); Carbonetti *et al., Abstr. Annu. Meet. Am. Soc. Microbiol.* 95:295 (1995); Sebo *et al., Infect. Immun.* 63:3851-3857 (1995); Klimpel *et al., PNAS U.S.A.* 89:10277-10281 (1992); and Novak *et al., J. Biol. Chem.* 267:17186-17193 1992)).

Such subsequences can be used to translocate ZFPs across a cell membrane. ZFPs can be conveniently fused to or derivatized with such sequences. Typically, the translocation sequence is provided as part of a fusion protein. Optionally, a linker can be used to link the ZFP and the translocation sequence. Any suitable linker can be used, e.g., a peptide linker.

The ZFP can also be introduced into an animal cell, preferably a mammalian cell, via a liposomes and liposome derivatives such as immunoliposomes. The term "liposome" refers to vesicles comprised of one or more concentrically ordered lipid bilayers, which encapsulate an aqueous phase. The aqueous phase typically contains the compound to be delivered to the cell, i.e., a ZFP.

The liposome fuses with the plasma membrane, thereby releasing the drug into the cytosol. Alternatively, the liposome is phagocytosed or taken up by the cell in a transport vesicle. Once in the endosome or phagosome, the liposome either degrades or fuses with the membrane of the transport vesicle and releases its contents.

In current methods of drug delivery via liposomes, the liposome ultimately becomes permeable and releases the encapsulated compound (in this case, a ZFP) at the target tissue or cell. For systemic or tissue specific delivery, this can be accomplished, for example, in a passive manner wherein the liposome bilayer degrades over time through the action of various agents in the body. Alternatively, active drug release

administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention, as described below (*see, e.g., Remington's Pharmaceutical Sciences*, 17th ed., 1989).

Delivery vehicles for ZFPs

An important factor in the administration of polypeptide compounds, such as the ZFPs, is ensuring that the polypeptide has the ability to traverse the plasma membrane of a cell, or the membrane of an intra-cellular compartment such as the nucleus. Cellular membranes are composed of lipid-protein bilayers that are freely permeable to small, nonionic lipophilic compounds and are inherently impermeable to polar compounds, macromolecules, and therapeutic or diagnostic agents. However, proteins and other compounds such as liposomes have been described, which have the ability to translocate polypeptides such as ZFPs across a cell membrane.

For example, "membrane translocation polypeptides" have amphiphilic or hydrophobic amino acid subsequences that have the ability to act as membrane-translocating carriers. In one embodiment, homeodomain proteins have the ability to translocate across cell membranes. The shortest internalizable peptide of a homeodomain protein, Antennapedia, was found to be the third helix of the protein, from amino acid position 43 to 58 (*see, e.g.*, Prochiantz, *Current Opinion in Neurobiology* 6:629-634 (1996)). Another subsequence, the h (hydrophobic) domain of signal peptides, was found to have similar cell membrane translocation characteristics (*see, e.g.*, Lin *et al., J. Biol. Chem.* 270:1 4255-14258 (1995)).

Examples of peptide sequences which can be linked to a ZFP of the invention, for facilitating uptake of ZFP into cells, include, but are not limited to: an 11 animo acid peptide of the tat protein of HIV; a 20 residue peptide sequence which corresponds to amino acids 84-103 of the p16 protein (*see* Fahraeus *et al., Current Biology* 6:84 (1996)); the third helix of the 60-amino acid long homeodomain of Antennapedia (Derossi *et al., J. Biol. Chem.* 269:10444 (1994)); the h region of a signal peptide such as the Kaposi fibroblast growth factor (K-FGF) h region (Lin *et al., supra*); or the VP22 translocation domain from HSV (Elliot & O'Hare, *Cell* 88:223-233 (1997)). Other suitable chemical moieties that provide enhanced cellular uptake may also be chemically linked to ZFPs.

from an individual patient (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

Ex vivo cell transfection for diagnostics, research, or for gene therapy (e.g., via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In a preferred embodiment, cells are isolated from the subject organism, transfected with a ZFP nucleic acid (gene or cDNA), and re-infused back into the subject organism (e.g., patient). Various cell types suitable for *ex vivo* transfection are well known to those of skill in the art (*see, e.g.*, Freshney *et al., Culture of Animal Cells, A Manual of Basic Technique* (3rd ed. 1994)) and the references cited therein for a discussion of how to isolate and culture cells from patients).

In one embodiment, stem cells are used in *ex vivo* procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types *in vitro*, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ cells *in vitro* into clinically important immune cell types using cytokines such a GM-CSF, IFN-γ and TNF-α are known (*see* Inaba *et al., J. Exp. Med.* 176:1693-1702 (1992)).

Stem cells are isolated for transduction and differentiation using known methods. For example, stem cells are isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panB cells), GR-1 (granulocytes), and Iad (differentiated antigen presenting cells) (*see* Inaba *et al., J. Exp. Med.* 176:1693-1702 (1992)).

Vectors (e.g., retroviruses, adenoviruses, liposomes, etc.) containing therapeutic ZFP nucleic acids can be also administered directly to the organism for transduction of cells *in vivo*. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and mor effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to

cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are supplied in *trans* by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. A viral vector is typically modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the viruses outer surface. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han *et al.*, *PNAS* 92:9747-9751 (1995), reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, filamentous phage can be engineered to display antibody fragments (e.g., FAB or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

Gene therapy vectors can be delivered *in vivo* by administration to an individual patient, typically by systemic administration (e.g., intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted

system. All of these viral vectors utilize approaches that involve complementation of defective vectors by genes inserted into helper cell lines to generate the transducing agent.

pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials (Dunbar *et al., Blood* 85:3048-305 (1995); Kohn *et al., Nat. Med.* 1:1017-102 (1995); Malech *et al., PNAS* 94:22 12133-12138 (1997)). PA317/pLASN was the first therapeutic vector used in a gene therapy trial. (Blaese *et al., Science* 270:475-480 (1995)). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors. (Ellem *et al., Immunol Immunother.* 44(1):10-20 (1997); Dranoff *et al., Hum. Gene Ther.* 1:111-2 (1997).

Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. All vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. (Wagner *et al., Lancet* 351:9117 1702-3 (1998), Kearns *et al., Gene Ther.* 9:748-55 (1996)).

Replication-deficient recombinant adenoviral vectors (Ad) are predominantly used for colon cancer gene therapy, because they can be produced at high titer and they readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the Ad E1a, E1b, and E3 genes; subsequently the replication defector vector is propagated in human 293 cells that supply deleted gene function in trans. Ad vectors can transduce multiply types of tissues *in vivo*, including nondividing, differentiated cells such as those found in the liver, kidney and muscle system tissues. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection (Sterman *et al., Hum. Gene Ther.* 7:1083-9 (1998)). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker *et al., Infection* 24:1 5-10 (1996); Sterman *et al., Hum. Gene Ther.* 9:7 1083-1089 (1998); Welsh *et al., Hum. Gene Ther.* 2:205-18 (1995); Alvarez *et al., Hum. Gene Ther.* 5:597-613 (1997); Topf *et al., Gene Ther.* 5:507-513 (1998); Sterman *et al., Hum. Gene Ther.* 7:1083-1089 (1998).

Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2

the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vector that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of *cis*-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum *cis*-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (*see, e.g.,* Buchscher *et al., J. Virol.* 66:2731-2739 (1992); Johann *et al., J. Virol.* 66:1635-1640 (1992); Sommerfelt *et al., Virol.* 176:58-59 (1990); Wilson *et al., J. Virol.* 63:2374-2378 (1989); Miller *et al., J. Virol.* 65:2220-2224 (1991); PCT/US94/05700).

In applications where transient expression of the ZFP is preferred, adenoviral based systems are typically used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, e.g., in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (*see, e.g.*, West *et al., Virology* 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin *et al., Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, *et al., Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulski *et al., J. Virol.* 63:03822-3828 (1989).

In particular, at least six viral vector approaches are currently available for gene transfer in clinical trials, with retroviral vectors by far the most frequently used

(1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada *et al.*, in *Current Topics in Microbiology and Immunology* Doerfler and Böhm (eds) (1995); and Yu *et al.*, *Gene Therapy* 1:13-26 (1994).

Methods of non-viral delivery of nucleic acids encoding engineered ZFPs include lipofection, microinjection, ballistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., US 5,049,386, US 4,946,787; and US 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024. Delivery can be to cells (*ex vivo* administration) or target tissues (*in vivo* administration).

The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (*see, e.g.*, Crystal, *Science* 270:404-410 (1995); Blaese *et al.*, *Cancer Gene Ther.* 2:291-297 (1995); Behr *et al.*, *Bioconjugate Chem.* 5:382-389 (1994); Remy *et al.*, *Bioconjugate Chem.* 5:647-654 (1994); Gao *et al.*, *Gene Therapy* 2:710-722 (1995); Ahmad *et al.*, *Cancer Res.* 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding engineered ZFP take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (*in vivo*) or they can be used to treat cells *in vitro* and the modified cells are administered to patients (*ex vivo*). Conventional viral based systems for the delivery of ZFPs could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Viral vectors are currently the most efficient and versatile method of gene transfer in target cells and tissues. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of

Another example of a preferred assay format useful for monitoring ZFP regulation of endogenous gene expression is performed *in vivo*. This assay is particularly useful for examining ZFPs that inhibit expression of tumor promoting genes, genes involved in tumor support, such as neovascularization (e.g., VEGF), or that activate tumor suppressor genes such as p53. In this assay, cultured tumor cells expressing the ZFP of choice are injected subcutaneously into an immune compromised mouse such as an athymic mouse, an irradiated mouse, or a SCID mouse. After a suitable length of time, preferably 4-8 weeks, tumor growth is measured, e.g., by volume or by its two largest dimensions, and compared to the control. Tumors that have statistically significant reduction (using, e.g., Student's T test) are said to have inhibited growth. Alternatively, the extent of tumor neovascularization can also be measured. Immunoassays using endothelial cell specific antibodies are used to stain for vascularization of the tumor and the number of vessels in the tumor. Tumors that have a statistically significant reduction in the number of vessels (using, e.g., Student's T test) are said to have inhibited neovascularization.

Transgenic and non-transgenic animals are also used as a preferred embodiment for examining regulation of endogenous gene expression *in vivo*. Transgenic animals typically express the ZFP of choice. Alternatively, animals that transiently express the ZFP of choice, or to which the ZFP has been administered in a delivery vehicle, can be used. Regulation of endogenous gene expression is tested using any one of the assays described herein.

Nucleic acids encoding ZFPs and gene therapy

Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding engineered ZFP in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding ZFPs to cells *in vitro*. Preferably, the nucleic acids encoding ZFPs are administered for *in vivo* or *ex vivo* gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, *Science* 256:808-813 (1992); Nabel & Felgner, *TIBTECH* 11:211-217 (1993); Mitani & Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175

Modulation of gene expression is tested using one of the *in vitro* or *in vivo* assays described herein. Samples or assays are treated with a ZFP and compared to control samples without the test compound, to examine the extent of modulation. As described above, for regulation of endogenous gene expression, the ZFP typically has a K_d of 200 nM or less, more preferably 100 nM or less, more preferably 50 nM, most preferably 25 nM or less.

The effects of the ZFPs can be measured by examining any of the parameters described above. Any suitable gene expression, phenotypic, or physiological change can be used to assess the influence of a ZFP. When the functional consequences are determined using intact cells or animals, one can also measure a variety of effects such as tumor growth, neovascularization, hormone release, transcriptional changes to both known and uncharacterized genetic markers (e.g., northern blots or oligonucleotide array studies), changes in cell metabolism such as cell growth or pH changes, and changes in intracellular second messengers such as cGMP.

Preferred assays for ZFP regulation of endogenous gene expression can be performed *in vitro*. In one preferred *in vitro* assay format, ZFP regulation of endogenous gene expression in cultured cells is measured by examining protein production using an ELISA assay (*see* Examples VI and VII). The test sample is compared to control cells treated with an empty vector or an unrelated ZFP that is targeted to another gene.

In another embodiment, ZFP regulation of endogenous gene expression is determined *in vitro* by measuring the level of target gene mRNA expression. The level of gene expression is measured using amplification, e.g., using PCR, LCR, or hybridization assays, e.g., northern hybridization, RNase protection, dot blotting. RNase protection is used in one embodiment (*see* Example VIII and Figure 10). The level of protein or mRNA is detected using directly or indirectly labeled detection agents, e.g., fluorescently or radioactively labeled nucleic acids, radioactively or enzymatically labeled antibodies, and the like, as described herein.

Alternatively, a reporter gene system can be devised using the target gene promoter operably linked to a reporter gene such as luciferase, green fluorescent protein, CAT, or β-gal. The reporter construct is typically co-transfected into a cultured cell. After treatment with the ZFP of choice, the amount of reporter gene transcription, translation, or activity is measured according to standard techniques known to those of skill in the art.

to standard techniques (*see, e.g.,* Morrison, *J. Bact.* 132:349-351 (1977); Clark-Curtiss & Curtiss, *Methods in Enzymology* 101:347-362 (Wu *et al.*, eds, 1983).

Any of the well known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see, e.g.,* Sambrook *et al.*, *supra*). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the protein of choice.

Assays for determining regulation of gene expression by ZFPs

A variety of assays can be used to determine the level of gene expression regulation by ZFPs. The activity of a particular ZFP can be assessed using a variety of *in vitro* and *in vivo* assays, by measuring, e.g., protein or mRNA levels, product levels, enzyme activity, tumor growth; transcriptional activation or repression of a reporter gene; second messenger levels (e.g., cGMP, cAMP, IP3, DAG, Ca^{2+}); cytokine and hormone production levels; and neovascularization, using, e.g., immunoassays (e.g., ELISA and immunohistochemical assays with antibodies), hybridization assays (e.g., RNase protection, northerns, *in situ* hybridization, oligonucleotide array studies), colorimetric assays, amplification assays, enzyme activity assays, tumor growth assays, phenotypic assays, and the like.

ZFPs are typically first tested for activity *in vitro* using cultured cells, e.g., 293 cells, CHO cells, VERO cells, BHK cells, HeLa cells, COS cells, and the like. Preferably, human cells are used. The ZFP is often first tested using a transient expression system with a reporter gene, and then regulation of the target endogenous gene is tested in cells and in animals, both *in vivo* and *ex vivo*. The ZFP can be recombinantly expressed in a cell, recombinantly expressed in cells transplanted into an animal, or recombinantly expressed in a transgenic animal, as well as administered as a protein to an animal or cell using delivery vehicles described below. The cells can be immobilized, be in solution, be injected into an animal, or be naturally occurring in a transgenic or non-transgenic animal.

The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the ZFP, e.g., expression in plants, animals, bacteria, fungus, protozoa etc. (see expression vectors described below and in the Example section). Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and commercially available fusion expression systems such as GST and LacZ. A preferred fusion protein is the maltose binding protein, "MBP." Such fusion proteins are used for purification of the ZFP. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, for monitoring expression, and for monitoring cellular and subcellular localization, e.g., c-myc or FLAG.

Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High yield expression systems are also suitable, such as using a baculovirus vector in insect cells, with a ZFP encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of recombinant sequences.

Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of protein, which are then purified using standard techniques (*see, e.g.,* Colley *et al., J. Biol. Chem.* 264:17619-17622 (1989); *Guide to Protein Purification, in Methods in Enzymology*, vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according

plant cell, animal cell, preferably a mammalian cell or a human cell, fungal cell, bacterial cell, or protozoal cell.

To obtain expression of a cloned gene or nucleic acid, a ZFP is typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, e.g., in Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994). Bacterial expression systems for expressing the ZFP are available in, e.g., *E. coli*, *Bacillus sp.*, and *Salmonella* (Palva *et al.*, *Gene* 22:229-235 (1983)). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available.

The promoter used to direct expression of a ZFP nucleic acid depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification of ZFP. In contrast, when a ZFP is administered *in vivo* for gene regulation, either a constitutive or an inducible promoter is used, depending on the particular use of the ZFP. In addition, a preferred promoter for administration of a ZFP can be a weak promoter, such as HSV TK or a promoter having similar activity. The promoter typically can also include elements that are responsive to transactivation, e.g., hypoxia response elements, Gal4 response elements, lac repressor response element, and small molecule control systems such as tet-regulated systems and the RU-486 system (*see, e.g.*, Gossen & Bujard, *PNAS* 89:5547 (1992); Oligino *et al.*, *Gene Ther.* 5:491-496 (1998); Wang *et al.*, *Gene Ther.* 4:432-441 (1997); Neering *et al.*, *Blood* 88:1147-1155 (1996); and Rendahl *et al.*, *Nat. Biotechnol.* 16:757-761 (1998)).

In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter operably linked, e.g., to the nucleic acid sequence encoding the ZFP, and signals required, e.g., for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding sites, or translation termination. Additional elements of the cassette may include, e.g., enhancers, and heterologous spliced intronic signals.

polypeptide sequences, such as polyglycine sequences of between about 5 and 200 amino acids. Preferred linkers are typically flexible amino acid subsequences which are synthesized as part of a recombinant fusion protein. For example, in one embodiment, the linker DGGGS is used to link two ZFPs. In another embodiment, the flexible linker linking two ZFPs is an amino acid subsequence comprising the sequence TGEKP (*see, e.g.*, Liu *et al.*, *PNAS* 5525-5530 (1997)). In another embodiment, the linker LRQKDGERP is used to link two ZFPs. In another embodiment, the following linkers are used to link two ZFPs: GGRR (Pomerantz *et al.* 1995, *supra*), $(G4S)_n$ (Kim *et al.*, *PNAS* 93, 1156-1160 (1996); and GGRRGGGS; LRQRDGERP; LRQKDGGGSERP; LRQKD$(G3S)_2$ ERP. Alternatively, flexible linkers can be rationally designed using computer program capable of modeling both DNA-binding sites and the peptides themselves (Desjarlais & Berg, *PNAS* 90:2256-2260 (1993), *PNAS* 91:11099-11103 (1994) or by phage display methods.

In other embodiments, a chemical linker is used to connect synthetically or recombinantly produced domain sequences. Such flexible linkers are known to persons of skill in the art. For example, poly(ethylene glycol) linkers are available from Shearwater Polymers, Inc. Huntsville, Alabama. These linkers optionally have amide linkages, sulfhydryl linkages, or heterofunctional linkages. In addition to covalent linkage of ZFPs to regulatory domains, non-covalent methods can be used to produce molecules with ZFPs associated with regulatory domains.

In addition to regulatory domains, often the ZFP is expressed as a fusion protein such as maltose binding protein ("MBP"), glutathione S transferase (GST), hexahistidine, c-myc, and the FLAG epitope, for ease of purification, monitoring expression, or monitoring cellular and subcellular localization.

Expression vectors for nucleic acids encoding ZFP

The nucleic acid encoding the ZFP of choice is typically cloned into intermediate vectors for transformation into prokaryotic or eukaryotic cells for replication and/or expression, e.g., for determination of K_d. Intermediate vectors are typically prokaryote vectors, e.g., plasmids, or shuttle vectors, or insect vectors, for storage or manipulation of the nucleic acid encoding ZFP or production of protein. The nucleic acid encoding a ZFP is also typically cloned into an expression vector, for administration to a

example, Vos, *Curr. Opin. Cell Biol.* 4:385-95 (1992); Sancar, *Ann. Rev. Genet.* 29:69-105 (1995); Lehmann, *Genet. Eng.* 17:1-19 (1995); and Wood, *Ann. Rev. Biochem.* 65:135-67 (1996). DNA rearrangement enzymes and their associated factors and modifiers can also be used as regulatory domains (*see, e.g.,* Gangloff *et al., Experientia* 50:261-9 (1994); Sadowski, *FASEB J.* 7:760-7 (1993)).

Similarly, regulatory domains can be derived from DNA modifying enzymes (e.g., DNA methyltransferases, topoisomerases, helicases, ligases, kinases, phosphatases, polymerases) and their associated factors and modifiers. Helicases are reviewed in Matson *et al., Bioessays*, 16:13-22 (1994), and methyltransferases are described in Cheng, *Curr. Opin. Struct. Biol.* 5:4-10 (1995). Chromatin associated proteins and their modifiers (e.g., kinases, acetylases and deacetylases), such as histone deacetylase (Wolffe, *Science* 272:371-2 (1996)) are also useful as domains for addition to the ZFP of choice. In one preferred embodiment, the regulatory domain is a DNA methyl transferase that acts as a transcriptional repressor (*see, e.g.,* Van den Wyngaert *et al., FEBS Lett.* 426:283-289 (1998); Flynn *et al., J. Mol. Biol.* 279:101-116 (1998); Okano *et al., Nucleic Acids Res.* 26:2536-2540 (1998); and Zardo & Caiafa, *J. Biol. Chem.* 273:16517-16520 (1998)). In another preferred embodiment, endonucleases such as Fok1 are used as transcriptional repressors, which act via gene cleavage (*see, e.g.,* WO95/09233; and PCT/US94/01201).

Factors that control chromatin and DNA structure, movement and localization and their associated factors and modifiers; factors derived from microbes (e.g., prokaryotes, eukaryotes and virus) and factors that associate with or modify them can also be used to obtain chimeric proteins. In one embodiment, recombinases and integrases are used as regulatory domains. In one embodiment, histone acetyltransferase is used as a transcriptional activator (*see, e.g.,* Jin & Scotto, *Mol. Cell. Biol.* 18:4377-4384 (1998); Wolffe, *Science* 272:371-372 (1996); Taunton *et al., Science* 272:408-411 (1996); and Hassig *et al., PNAS* 95:3519-3524 (1998)). In another embodiment, histone deacetylase is used as a transcriptional repressor (*see, e.g.,* Jin & Scotto, *Mol. Cell. Biol.* 18:4377-4384 (1998); Syntichaki & Thireos, *J. Biol. Chem.* 273:24414-24419 (1998); Sakaguchi *et al., Genes Dev.* 12:2831-2841 (1998); and Martinez *et al., J. Biol. Chem.* 273:23781-23785 (1998)).

Linker domains between polypeptide domains, e.g., between two ZFPs or between a ZFP and a regulatory domain, can be included. Such linkers are typically

In one embodiment, the HSV VP16 activation domain is used as a transcriptional activator (*see, e.g.*, Hagmann *et al., J. Virol.* 71:5952-5962 (1997)). Other preferred transcription factors that could supply activation domains include the VP64 activation domain (Seipel *et al., EMBO J.* 11:4961-4968 (1996)); nuclear hormone receptors (*see, e.g.*, Torchia *et al., Curr. Opin. Cell. Biol.* 10:373-383 (1998)); the p65 subunit of nuclear factor kappa B (Bitko & Barik, *J. Virol.* 72:5610-5618 (1998) and Doyle & Hunt, *Neuroreport* 8:2937-2942 (1997)); and EGR-1 (early growth response gene product-1; Yan *et al., PNAS* 95:8298-8303 (1998); and Liu *et al., Cancer Gene Ther.* 5:3-28 (1998)).

Kinases, phosphatases, and other proteins that modify polypeptides involved in gene regulation are also useful as regulatory domains for ZFPs. Such modifiers are often involved in switching on or off transcription mediated by, for example, hormones. Kinases involved in transcription regulation are reviewed in Davis, *Mol. Reprod. Dev.* 42:459-67 (1995), Jackson *et al., Adv. Second Messenger Phosphoprotein Res.* 28:279-86 (1993), and Boulikas, *Crit. Rev. Eukaryot. Gene Expr.* 5:1-77 (1995), while phosphatases are reviewed in, for example, Schonthal & Semin, *Cancer Biol.* 6:239-48 (1995). Nuclear tyrosine kinases are described in Wang, *Trends Biochem. Sci.* 19:373-6 (1994).

As described, useful domains can also be obtained from the gene products of oncogenes (e.g., myc, jun, fos, myb, max, mad, rel, ets, bcl, mos family members) and their associated factors and modifiers. Oncogenes are described in, for example, Cooper, *Oncogenes*, 2nd ed., The Jones and Bartlett Series in Biology, Boston, MA, Jones and Bartlett Publishers, 1995. The ets transcription factors are reviewed in Waslylk *et al., Eur. J. Biochem.* 211:7-18 (1993) and Crepieux *et al., Crit. Rev. Oncog.* 5:615-38 (1994). Myc oncogenes are reviewed in, for example, Ryan *et al., Biochem. J.* 314:713-21 (1996). The jun and fos transcription factors are described in, for example, *The Fos and Jun Families of Transcription Factors*, Angel & Herrlich, eds. (1994). The max oncogene is reviewed in Hurlin *et al.*, Cold Spring Harb. Symp. Quant. Biol. 59:109-16. The myb gene family is reviewed in Kanei-Ishii *et al., Curr. Top. Microbiol. Immunol.* 211:89-98 (1996). The mos family is reviewed in Yew *et al., Curr. Opin. Genet. Dev.* 3:19-25 (1993).

ZFPs can include regulatory domains obtained from DNA repair enzymes and their associated factors and modifiers. DNA repair systems are reviewed in, for

factors are known (*see, e.g., Science* 269:630 (1995)). Nuclear hormone receptor transcription factors are described in, for example, Rosen *et al., J. Med. Chem.* 38:4855-74 (1995). The C/EBP family of transcription factors are reviewed in Wedel *et al., Immunobiology* 193:171-85 (1995). Coactivators and co-repressors that mediate transcription regulation by nuclear hormone receptors are reviewed in, for example, Meier, *Eur. J. Endocrinol.* 134(2):158-9 (1996); Kaiser *et al., Trends Biochem. Sci.* 21:342-5 (1996); and Utley *et al., Nature* 394:498-502 (1998)). GATA transcription factors, which are involved in regulation of hematopoiesis, are described in, for example, Simon, *Nat. Genet.* 11:9-11 (1995); Weiss *et al., Exp. Hematol.* 23:99-107. TATA box binding protein (TBP) and its associated TAF polypeptides (which include TAF30, TAF55, TAF80, TAF110, TAF150, and TAF250) are described in Goodrich & Tjian, *Curr. Opin. Cell Biol.* 6:403-9 (1994) and Hurley, *Curr. Opin. Struct. Biol.* 6:69-75 (1996). The STAT family of transcription factors are reviewed in, for example, Barahmand-Pour *et al., Curr. Top. Microbiol. Immunol.* 211:121-8 (1996). Transcription factors involved in disease are reviewed in Aso *et al., J. Clin. Invest.* 97:1561-9 (1996).

In one embodiment, the KRAB repression domain from the human KOX-1 protein is used as a transcriptional repressor (Thiesen *et al., New Biologist* 2:363-374 (1990); Margolin *et al., PNAS* 91:4509-4513 (1994); Pengue *et al., Nucl. Acids Res.* 22:2908-2914 (1994); Witzgall *et al., PNAS* 91:4514-4518 (1994); *see also* Example III)). In another embodiment, KAP-1, a KRAB co-repressor, is used with KRAB (Friedman *et al., Genes Dev.* 10:2067-2078 (1996)). Alternatively, KAP-1 can be used alone with a ZFP. Other preferred transcription factors and transcription factor domains that act as transcriptional repressors include MAD (*see, e.g.*, Sommer *et al., J. Biol. Chem.* 273:6632-6642 (1998); Gupta *et al., Oncogene* 16:1149-1159 (1998); Queva *et al., Oncogene* 16:967-977 (1998); Larsson *et al., Oncogene* 15:737-748 (1997); Laherty *et al., Cell* 89:349-356 (1997); and Cultraro *et al., Mol Cell. Biol.* 17:2353-2359 (19977)); FKHR (forkhead in rhapdosarcoma gene; Ginsberg *et al., Cancer Res.* 15:3542-3546 (1998); Epstein *et al., Mol. Cell. Biol.* 18:4118-4130 (1998)); EGR-1 (early growth response gene product-1; Yan *et al., PNAS* 95:8298-8303 (1998); and Liu *et al., Cancer Gene Ther.* 5:3-28 (1998)); the ets2 repressor factor repressor domain (ERD; Sgouras *et al., EMBO J.* 14:4781-4793 ((19095)); and the MAD smSIN3 interaction domain (SID; Ayer *et al., Mol. Cell. Biol.* 16:5772-5781 (1996)).

then repeated as many times as is necessary to sufficiently enrich the phage pool for tight binders such that these may be identified using sequencing and/or screening methods.

Regulatory domains

The ZFPs of the invention can optionally be associated with regulatory domains for modulation of gene expression. The ZFP can be covalently or non-covalently associated with one or more regulatory domains, alternatively two or more regulatory domains, with the two or more domains being two copies of the same domain, or two different domains. The regulatory domains can be covalently linked to the ZFP, e.g., via an amino acid linker, as part of a fusion protein. The ZFPs can also be associated with a regulatory domain via a non-covalent dimerization domain, e.g., a leucine zipper, a STAT protein N terminal domain, or an FK506 binding protein (*see, e.g.*, O'Shea, *Science* 254: 539 (1991), Barahmand-Pour *et al.*, *Curr. Top. Microbiol. Immunol.* 211:121-128 (1996); Klemm *et al.*, *Annu. Rev. Immunol.* 16:569-592 (1998); Klemm *et al.*, *Annu. Rev. Immunol.* 16:569-592 (1998); Ho *et al.*, *Nature* 382:822-826 (1996); and Pomeranz *et al.*, *Biochem.* 37:965 (1998)). The regulatory domain can be associated with the ZFP at any suitable position, including the C- or N-terminus of the ZFP.

Common regulatory domains for addition to the ZFP include, e.g., effector domains from transcription factors (activators, repressors, co-activators, co-repressors), silencers, nuclear hormone receptors, oncogene transcription factors (e.g., myc, jun, fos, myb, max, mad, rel, ets, bcl, mos family members etc.); DNA repair enzymes and their associated factors and modifiers; DNA rearrangement enzymes and their associated factors and modifiers; chromatin associated proteins and their modifiers (e.g., kinases, acetylases and deacetylases); and DNA modifying enzymes (e.g., methyltransferases, topoisomerases, helicases, ligases, kinases, phosphatases, polymerases, endonucleases) and their associated factors and modifiers.

Transcription factor polypeptides from which one can obtain a regulatory domain include those that are involved in regulated and basal transcription. Such polypeptides include transcription factors, their effector domains, coactivators, silencers, nuclear hormone receptors (*see, e.g.*, Goodrich *et al.*, *Cell* 84:825-30 (1996) for a review of proteins and nucleic acid elements involved in transcription; transcription factors in general are reviewed in Barnes & Adcock, *Clin. Exp. Allergy* 25 Suppl. 2:46-9 (1995) and Roeder, *Methods Enzymol.* 273:165-71 (1996)). Databases dedicated to transcription

Similar assays can also include determining active fractions in the protein preparations. Active fractions are determined by stoichiometric gel shifts where proteins are titrated against a high concentration of target DNA. Titrations are done at 100, 50, and 25% of target (usually at micromolar levels).

In another embodiment, phage display libraries can be used to select ZFPs with high affinity to the selected target site. This method differs fundamentally from direct design in that it involves the generation of diverse libraries of mutagenized ZFPs, followed by the isolation of proteins with desired DNA-binding properties using affinity selection methods. To use this method, the experimenter typically proceeds as follows.

First, a gene for a ZFP is mutagenized to introduce diversity into regions important for binding specificity and/or affinity. In a typical application, this is accomplished via randomization of a single finger at positions –1, +2, +3, and +6, and perhaps accessory positions such as +1, +5, +8, or +10.

Next, the mutagenized gene is cloned into a phage or phagemid vector as a fusion with, e.g., gene III of filamentous phage, which encodes the coat protein pIII. The zinc finger gene is inserted between segments of gene III encoding the membrane export signal peptide and the remainder of pIII, so that the ZFP is expressed as an amino-terminal fusion with pIII in the mature, processed protein. When using phagemid vectors, the mutagenized zinc finger gene may also be fused to a truncated version of gene III encoding, minimally, the C-terminal region required for assembly of pIII into the phage particle.

The resultant vector library is transformed into *E. coli* and used to produce filamentous phage which express variant ZFPs on their surface as fusions with the coat protein pIII (if a phagemid vector is used, then the this step requires superinfection with helper phage). The phage library is then incubated with target DNA site, and affinity selection methods are used to isolate phage which bind target with high affinity from bulk phage. Typically, the DNA target is immobilized on a solid support, which is then washed under conditions sufficient to remove all but the tightest binding phage. After washing, any phage remaining on the support are recovered via elution under conditions which totally disrupt zinc finger-DNA binding.

Recovered phage are used to infect fresh *E. coli*, which is then amplified and used to produce a new batch of phage particles. The binding and recovery steps are

protein can be obtained by induction with IPTG since the MBP-ZFP fusion in the pMal-c2 expression plasmid is under the control of the IPTG inducible tac promoter (NEB). Bacteria containing the MBP-ZFP fusion plasmids are inoculated in to 2xYT medium containing 10μM $ZnCl_2$, 0.02% glucose, plus 50 μg/ml ampicillin and shaken at 37°C. At mid-exponential growth IPTG is added to 0.3 mM and the cultures are allowed to shake. After 3 hours the bacteria are harvested by centrifugation, disrupted by sonication, and then insoluble material is removed by centrifugation. The MBP-ZFP proteins are captured on an amylose-bound resin, washed extensively with buffer containing 20 mM Tris-HCl (pH 7.5), 200 mM NaCl, 5 mM DTT and 50 μM $ZnCl_2$, then eluted with maltose in essentially the same buffer (purification is based on a standard protocol from NEB). Purified proteins are quantitated and stored for biochemical analysis.

The biochemical properties of the purified proteins, e.g., K_d, can be characterized by any suitable assay. In one embodiment, K_d is characterized via electrophoretic mobility shift assays ("EMSA") (Buratowski & Chodosh, *in Current Protocols in Molecular Biology* pp. 12.2.1-12.2.7 (Ausubel ed., 1996); *see also* U.S. Patent No. 5,789,538, USSN 09/229,007, filed January 12, 1999, herein incorporated by reference, and Example I). Affinity is measured by titrating purified protein against a low fixed amount of labeled double-stranded oligonucleotide target. The target comprises the natural binding site sequence (9 or 18 bp) flanked by the 3 bp found in the natural sequence. External to the binding site plus flanking sequence is a constant sequence. The annealed oligonucleotide targets possess a 1 bp 5' overhang which allows for efficient labeling of the target with T4 phage polynucleotide kinase. For the assay the target is added at a concentration of 40 nM or lower (the actual concentration is kept at least 10-fold lower than the lowest protein dilution) and the reaction is allowed to equilibrate for at least 45 min. In addition the reaction mixture also contains 10 mM Tris (pH 7.5), 100 mM KCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 5 mM DTT, 10% glycerol, 0.02% BSA (poly (dIdC) or (dAdT) (Pharmacia) can also added at 10-100 μg/μl).

The equilibrated reactions are loaded onto a 10% polyacrylamide gel, which has been pre-run for 45 min in Tris/glycine buffer, then bound and unbound labeled target is resolved be electrophoresis at 150V (alternatively, 10-20% gradient Tris-HCl gels, containing a 4% polyacrylamide stacker, can be used). The dried gels are visualized by autoradiography or phosphoroimaging and the apparent K_d is determined by calculating the protein concentration that gives half-maximal binding.

filled in by high-fidelity thermostable polymerase, the combination of Taq and Pfu polymerases also suffices. In the second phase of construction, the zinc finger template is amplified by external primers designed to incorporate restriction sites at either end for cloning into a shuttle vector or directly into an expression vector.

An alternative method of cloning the newly designed DNA-binding proteins relies on annealing complementary oligonucleotides encoding the specific regions of the desired ZFP. This particular application requires that the oligonucleotides be phosphorylated prior to the final ligation step. This is usually performed before setting up the annealing reactions, but kinasing can also occur post-annealing. In brief, the "universal" oligonucleotides encoding the constant regions of the proteins (oligos 1, 2 and 3 of above) are annealed with their complementary oligonucleotides. Additionally, the "specific" oligonucleotides encoding the finger recognition helices are annealed with their respective complementary oligonucleotides. These complementary oligos are designed to fill in the region which was previously filled in by polymerase in the protocol described above. The complementary oligos to the common oligos 1 and finger 3 are engineered to leave overhanging sequences specific for the restriction sites used in cloning into the vector of choice. The second assembly protocol differs from the initial protocol in the following aspects: the "scaffold" encoding the newly designed ZFP is composed entirely of synthetic DNA thereby eliminating the polymerase fill-in step, additionally the fragment to be cloned into the vector does not require amplification. Lastly, the design of leaving sequence-specific overhangs eliminates the need for restriction enzyme digests of the inserting fragment.

The resulting fragment encoding the newly designed ZFP is ligated into an expression vector. Expression vectors that are commonly utilized include, but are not limited to, a modified pMAL-c2 bacterial expression vector (New England BioLabs, "NEB") or a eukaryotic expression vector, pcDNA (Promega).

Any suitable method of protein purification known to those of skill in the art can be used to purify ZFPs of the invention (*see* Ausubel, *supra*, Sambrook, *supra*). In addition, any suitable host can be used, e.g., bacterial cells, insect cells, yeast cells, mammalian cells, and the like.

In one embodiment, expression of the ZFP fused to a maltose binding protein (MBP-ZFP) in bacterial strain JM109 allows for straightforward purification through an amylose column (NEB). High expression levels of the zinc finger chimeric

specificity for a given target by an empirical process such as phage display. In some such methods, each component finger of a ZFP is designed or selected independently of other component fingers. For example, each finger can be obtained from a different preexisting ZFP or each finger can be subject to separate randomization and selection.

Once a ZFP has been selected, designed, or otherwise provided to a given target segment, the ZFP or the DNA encoding it are synthesized. Exemplary methods for synthesizing and expressing DNA encoding zinc finger proteins are described below. The ZFP or a polynucleotide encoding it can then be used for modulation of expression, or analysis of the target gene containing the target site to which the ZFP binds.

Expression and purification of ZFPs

ZFP polypeptides and nucleic acids can be made using routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)). In addition, essentially any nucleic acid can be custom ordered from any of a variety of commercial sources. Similarly, peptides and antibodies can be custom ordered from any of a variety of commercial sources.

Two alternative methods are typically used to create the coding sequences required to express newly designed DNA-binding peptides. One protocol is a PCR-based assembly procedure that utilizes six overlapping oligonucleotides (Figure 1). Three oligonucleotides (oligos 1, 3, and 5 in Figure 1) correspond to "universal" sequences that encode portions of the DNA-binding domain between the recognition helices. These oligonucleotides remain constant for all zinc finger constructs. The other three "specific" oligonucleotides (oligos 2, 4, and 6 in Figure 1) are designed to encode the recognition helices. These oligonucleotides contain substitutions primarily at positions -1, 2, 3 and 6 on the recognition helices making them specific for each of the different DNA-binding domains.

The PCR synthesis is carried out in two steps. First, a double stranded DNA template is created by combining the six oligonucleotides (three universal, three specific) in a four cycle PCR reaction with a low temperature annealing step, thereby annealing the oligonucleotides to form a DNA "scaffold." The gaps in the scaffold are

In a variation, the methods of the invention identify first and second target segments, each independently conforming to the above formula. The two target segments in such methods are constrained to be adjacent or proximate (i.e., within about 0-5 bases) of each other in the target gene. The strategy underlying selection of proximate target segments is to allow the design of a ZFP formed by linkage of two component ZFPs specific for the first and second target segments respectively. These principles can be extended to select target sites to be bound by ZFPs with any number of component fingers. For example, a suitable target site for a nine finger protein would have three component segments, each conforming to the above formula.

The target sites identified by the above methods can be subject to further evaluation by other criteria or can be used directly for design or selection (if needed) and production of a ZFP specific for such a site. A further criteria for evaluating potential target sites is their proximity to particular regions within a gene. If a ZFP is to be used to repress a cellular gene on its own (i.e., without linking the ZFP to a repressing moiety), then the optimal location appears to be at, or within 50 bp upstream or downstream of the site of transcription initiation, to interfere with the formation of the transcription complex (Kim & Pabo, *J. Biol. Chem.* 272:29795-296800 (1997)) or compete for an essential enhancer binding protein. If, however, a ZFP is fused to a functional domain such as the KRAB repressor domain or the VP16 activator domain, the location of the binding site is considerably more flexible and can be outside known regulatory regions. For example, a KRAB domain can repress transcription at a promoter up to at least 3 kbp from where KRAB is bound (Margolin *et al.*, *PNAS* 91:4509-4513 (1994)). Thus, target sites can be selected that do not necessarily include or overlap segments of demonstrable biological significance with target genes, such as regulatory sequences. Other criteria for further evaluating target segments include the prior availability of ZFPs binding to such segments or related segments, and/or ease of designing new ZFPs to bind a given target segment.

After a target segment has been selected, a ZFP that binds to the segment can be provided by a variety of approaches. The simplest of approaches is to provide a precharacterized ZFP from an existing collection that is already known to bind to the target site. However, in many instances, such ZFPs do not exist. An alternative approach can also be used to design new ZFPs, which uses the information in a database of existing ZFPs and their respective binding affinities. A further approach is to design a ZFP based on substitution rules as discussed above. A still further alternative is to select a ZFP with

In the design of a ZFP with three fingers, a target site should be selected in which at least one finger of the protein, and optionally, two or all three fingers have the potential to bind a D-able site. Such can be achieved by selecting a target site from within a larger target gene having the formula 5'-NNx aNy bNzc-3', wherein

each of the sets (x, a), (y, b) and (z, c) is either (N, N) or (G, K);
at least one of (x, a), (y, b) and (z, c) is (G, K). and
N and K are IUPAC-IUB ambiguity codes

In other words, at least one of the three sets (x, a), (y, b) and (z, c) is the set (G, K), meaning that the first position of the set is G and the second position is G or T. Those of the three sets (if any) which are not (G, K) are (N, N), meaning that the first position of the set can be occupied by any nucleotide and the second position of the set can be occupied by any nucleotide. As an example, the set (x, a) can be (G, K) and the sets (y, b) and (z, c) can both be (N, N).

In the formula 5'-NNx aNy bNzc-3', the triplets of NNx aNy and bNzc represent the triplets of bases on the target strand bound by the three fingers in a ZFP. If only one of x, y and z is a G, and this G is followed by a K, the target site includes a single D-able subsite. For example, if only x is G, and a is K, the site reads 5'-**NNG KNy** bNzc-3' with the D-able subsite highlighted. If both x and y but not z are G, and a and b are K, then the target site has two overlapping D-able subsites as follows: 5'-**NNG** ***KNG*** ***K***Nz c-3', with one such site being represented in bold and the other in italics. If all three of x, y and z are G and a, b, and c are K, then the target segment includes three D-able subsites, as follows 5'**NNG** ***KNG*** <u>KNG K</u>3', the D-able subsites being represented by bold, italics and underline.

These methods thus work by selecting a target gene, and systematically searching within the possible subsequences of the gene for target sites conforming to the formula 5'-NNx aNy bNzc-3', as described above. In some such methods, every possible subsequence of 10 contiguous bases on either strand of a potential target gene is evaluated to determine whether it conforms to the above formula, and, if so, how many D-able sites are present. Typically, such a comparison is performed by computer, and a list of target sites conforming to the formula are output. Optionally, such target sites can be output in different subsets according to how many D-able sites are present.

94:5525-5530 (1997); Griesman & Pabo, *Science* 275:657-661 (1997); Desjarlais & Berg, *PNAS* 91:11-99-11103 (1994)).

In a preferred embodiment, copending application USSN 09/229,007, filed January 12, 1999 provides methods that select a target gene, and identify a target site within the gene containing one to six (or more) D-able sites (see definition below). Using these methods, a ZFP can then be synthesized that binds to the preselected site. These methods of target site selection are premised, in part, on the recognition that the presence of one or more D-able sites in a target segment confers the potential for higher binding affinity in a ZFP selected or designed to bind to that site relative to ZFPs that bind to target segments lacking D-able sites (see below). Experimental evidence supporting this insight is provided in Examples 2-9 of copending application USSN 09/229,007, filed January 12, 1999.

A D-able site or subsite is a region of a target site that allows an appropriately designed single zinc finger to bind to four bases rather than three of the target site. Such a zinc finger binds to a triplet of bases on one strand of a double-stranded target segment (target strand) and a fourth base on the other strand (*see* Figure 2 of copending application USSN 09/229,007, filed January 12, 1999. Binding of a single zinc finger to a four base target segment imposes constraints both on the sequence of the target strand and on the amino acid sequence of the zinc finger. The target site within the target strand should include the "D-able" site motif 5' NNGK 3', in which N and K are conventional IUPAC-IUB ambiguity codes. A zinc finger for binding to such a site should include an arginine residue at position -1 and an aspartic acid, (or less preferably a glutamic acid) at position +2. The arginine residues at position -1 interacts with the G residue in the D-able site. The aspartic acid (or glutamic acid) residue at position +2 of the zinc finger interacts with the opposite strand base complementary to the K base in the D-able site. It is the interaction between aspartic acid (symbol D) and the opposite strand base (fourth base) that confers the name D-able site. As is apparent from the D-able site formula, there are two subtypes of D-able sites: 5' NNGG 3' and 5' NNGT 3'. For the former site, the aspartic acid or glutamic acid at position +2 of a zinc finger interacts with a C in the opposite strand to the D-able site. In the latter site, the aspartic acid or glutamic acid at position +2 of a zinc finger interacts with an A in the opposite strand to the D-able site. In general, NNGG is preferred over NNGT.

The following eight groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Glycine (G);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
7) Serine (S), Threonine (T); and
8) Cysteine (C), Methionine (M)

(*see, e.g.*, Creighton, *Proteins* (1984)).

Design of ZFPs

The ZFPs of the invention are engineered to recognize a selected target site in the endogenous gene of choice. Typically, a backbone from any suitable C_2H_2 ZFP, such as SP-1, SP-1C, or ZIF268, is used as the scaffold for the engineered ZFP (*see, e.g.*, Jacobs, *EMBO J.* 11:4507 (1992); Desjarlais & Berg, *PNAS* 90:2256-2260 (1993)). A number of methods can then be used to design and select a ZFP with high affinity for its target (e.g., preferably with a K_d of less than about 25 nM). As described above, a ZFP can be designed or selected to bind to any suitable target site in the target endogenous gene, with high affinity. Co-pending patent application USSN 09/229,007, filed January 12, 1999, comprehensively describes methods for design, construction, and expression of ZFPs for selected target sites.

Any suitable method known in the art can be used to design and construct nucleic acids encoding ZFPs, e.g., phage display, random mutagenesis, combinatorial libraries, computer/rational design, affinity selection, PCR, cloning from cDNA or genomic libraries, synthetic construction and the like. (*see, e.g.*, U.S. Pat. No. 5,786,538; Wu *et al.*, *PNAS* 92:344-348 (1995); Jamieson *et al.*, *Biochemistry* 33:5689-5695 (1994); Rebar & Pabo, *Science* 263:671-673 (1994); Choo & Klug, *PNAS* 91:11163-11167 (1994); Choo & Klug, *PNAS* 91: 11168-11172 (1994); Desjarlais & Berg, *PNAS* 90:2256-2260 (1993); Desjarlais & Berg, *PNAS* 89:7345-7349 (1992); Pomerantz *et al.*, *Science* 267:93-96 (1995); Pomerantz *et al.*, *PNAS* 92:9752-9756 (1995); and Liu *et al.*, *PNAS*

backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

cells, or mammalian cells such as CHO, HeLa, 293, COS-1, and the like, e.g., cultured cells (*in vitro*), explants and primary cultures (*in vitro* and *ex vivo*), and cells *in vivo*.

"Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms also apply to amino acid polymers in which one or more amino acid residues is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymer.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription. As used herein, a promoter typically includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of certain RNA polymerase II type promoters, a TATA element, enhancer, CCAAT box, SP-1 site, etc. As used herein, a promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. The promoters often have an element that is responsive to transactivation by a DNA-binding moiety such as a polypeptide, e.g., a nuclear receptor, Gal4, the lac repressor and the like.

A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under certain environmental or developmental conditions.

A "weak promoter" refers to a promoter having about the same activity as a wild type herpes simplex virus ("HSV") thymidine kinase ("tk") promoter or a mutated HSV tk promoter, as described in Eisenberg & McKnight, *Mol. Cell. Biol.* 5:1940-1947 (1985).

The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell, and optionally integration or replication of the expression vector in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment, of viral or non-viral origin. Typically, the expression vector includes an "expression cassette," which comprises a nucleic acid to be transcribed operably linked to a promoter. The term expression vector also encompasses naked DNA operably linked to a promoter.

By "host cell" is meant a cell that contains a ZFP or an expression vector or nucleic acid encoding a ZFP. The host cell typically supports the replication or expression of the expression vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, fungal, protozoal, higher plant, insect, or amphibian

Activators and repressors include co-activators and co-repressors (*see, e.g.*, Utley *et al.*, *Nature* 394:498-502 (1998)).

A "regulatory domain" refers to a protein or a protein domain that has transcriptional modulation activity when tethered to a DNA binding domain, i.e., a ZFP. Typically, a regulatory domain is covalently or non-covalently linked to a ZFP to effect transcription modulation. Alternatively, a ZFP can act alone, without a regulatory domain, to effect transcription modulation.

The term "heterologous" is a relative term, which when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, a nucleic acid that is recombinantly produced typically has two or more sequences from unrelated genes synthetically arranged to make a new functional nucleic acid, e.g., a promoter from one source and a coding region from another source. The two nucleic acids are thus heterologous to each other in this context. When added to a cell, the recombinant nucleic acids would also be heterologous to the endogenous genes of the cell. Thus, in a chromosome, a heterologous nucleic acid would include an non-native (non-naturally occurring) nucleic acid that has integrated into the chromosome, or a non-native (non-naturally occurring) extrachromosomal nucleic acid. In contrast, a naturally translocated piece of chromosome would not be considered heterologous in the context of this patent application, as it comprises an endogenous nucleic acid sequence that is native to the mutated cell.

Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (e.g., a "fusion protein," where the two subsequences are encoded by a single nucleic acid sequence). *See, e.g.*, Ausubel, *supra*, for an introduction to recombinant techniques.

The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (naturally occurring) form of the cell or express a second copy of a native gene that is otherwise normally or abnormally expressed, under expressed or not expressed at all.

"Inhibition of gene expression that prevents gene activation" refers to the ability of a zinc finger protein to block or prevent binding of an activator molecule.

Modulation can be assayed by determining any parameter that is indirectly or directly affected by the expression of the target gene. Such parameters include, e.g., changes in RNA or protein levels, changes in protein activity, changes in product levels, changes in downstream gene expression, changes in reporter gene transcription (luciferase, CAT, β-galactosidase, β-glucuronidase, GFP (*see, e.g.,* Mistili & Spector, *Nature Biotechnology* 15:961-964 (1997)); changes in signal transduction, phosphorylation and dephosphorylation, receptor-ligand interactions, second messenger concentrations (e.g., cGMP, cAMP, IP3, and Ca^{2+}), cell growth, and neovascularization. These assays can be *in vitro*, *in vivo*, and *ex vivo*. Such functional effects can be measured by any means known to those skilled in the art, e.g., measurement of RNA or protein levels, measurement of RNA stability, identification of downstream or reporter gene expression, e.g., via chemiluminescence, fluorescence, colorimetric reactions, antibody binding, inducible markers, ligand binding assays; changes in intracellular second messengers such as cGMP and inositol triphosphate (IP3); changes in intracellular calcium levels; cytokine release, and the like.

To determine the level of gene expression modulation by a ZFP, cells contacted with ZFPs are compared to control cells, e.g., without the zinc finger protein or with a non-specific ZFP, to examine the extent of inhibition or activation. Control samples are assigned a relative gene expression activity value of 100%. Modulation/inhibition of gene expression is achieved when the gene expression activity value relative to the control is about 80%, preferably 50% (i.e., 0.5x the activity of the control), more preferably 25%, more preferably 5-0%. Modulation/activation of gene expression is achieved when the gene expression activity value relative to the control is 110% , more preferably 150% (i.e., 1.5x the activity of the control), more preferably 200-500%, more preferably 1000-2000% or more.

A "transcriptional activator" and a "transcriptional repressor" refer to proteins or effector domains of proteins that have the ability to modulate transcription, as described above. Such proteins include, e.g., transcription factors and co-factors (e.g., KRAB, MAD, ERD, SID, nuclear factor kappa B subunit p65, early growth response factor 1, and nuclear hormone receptors, VP16, VP64), endonucleases, integrases, recombinases, methyltransferases, histone acetyltransferases, histone deacetylases etc.

The phrase "RNA polymerase pause site" is described in Uptain *et al.*, *Annu. Rev. Biochem.* 66:117-172 (1997).

"Humanized" refers to a non-human polypeptide sequence that has been modified to minimize immunoreactivity in humans, typically by altering the amino acid sequence to mimic existing human sequences, without substantially altering the function of the polypeptide sequence (*see, e.g.*, Jones *et al.*, *Nature* 321:522-525 (1986), and published UK patent application No. 8707252). Backbone sequences for the ZFPs are preferably be selected from existing human C_2H_2 ZFPs (e.g., SP-1). Functional domains are preferably selected from existing human genes, (e.g., the activation domain from the p65 subunit of NF-κB). Where possible, the recognition helix sequences will be selected from the thousands of existing ZFP DNA recognition domains provided by sequencing the human genome. As much as possible, domains will be combined as units from the same existing proteins. All of these steps will minimize the introduction of new junctional epitopes in the chimeric ZFPs and render the engineered ZFPs less immunogenic.

"Administering" an expression vector, nucleic acid, ZFP, or a delivery vehicle to a cell comprises transducing, transfecting, electroporating, translocating, fusing, phagocytosing, shooting or ballistic methods, etc., i.e., any means by which a protein or nucleic acid can be transported across a cell membrane and preferably into the nucleus of a cell.

A "delivery vehicle" refers to a compound, e.g., a liposome, toxin, or a membrane translocation polypeptide, which is used to administer a ZFP. Delivery vehicles can also be used to administer nucleic acids encoding ZFPs, e.g., a lipid:nucleic acid complex, an expression vector, a virus, and the like.

The terms "modulating expression" "inhibiting expression" and "activating expression" of a gene refer to the ability of a ZFP to activate or inhibit transcription of a gene. Activation includes prevention of transcriptional inhibition (i.e., prevention of repression of gene expression) and inhibition includes prevention of transcriptional activation (i.e., prevention of gene activation).

"Activation of gene expression that prevents repression of gene expression" refers to the ability of a zinc finger protein to block or prevent binding of a repressor molecule.

using an electrophoretic mobility shift assay ("EMSA"), as described in Example I and on page 14 of the present specification. Unless an adjustment is made for ZFP purity or activity, the K_d calculations made using the method of Example I may result in an underestimate of the true K_d of a given ZFP. Preferably, the K_d of a ZFP used to modulate transcription of an endogenous cellular gene is less than about 100 nM, more preferably less than about 75 nM, more preferably less than about 50 nM, most preferably less than about 25 nM.

An "endogenous cellular gene" refers to a gene that is native to a cell, which is in its normal genomic and chromatin context, and which is not heterologous to the cell. Such cellular genes include, e.g., animal genes, plant genes, bacterial genes, protozoal genes, fungal genes, mitrochondrial genes, and chloroplastic genes.

An "endogenous gene" refers to a microbial or viral gene that is part of a naturally occurring microbial or viral genome in a microbially or virally infected cell. The microbial or viral genome can be extrachromosomal or integrated into the host chromosome. This term also encompasses endogenous cellular genes, as described above.

A "native chromatin environment" refers to the naturally occurring, structural relationship of genomic DNA (e.g., bacterial, animal, fungal, plant, protozoal, mitochondrial, and chloroplastic) and DNA-binding proteins (e.g., histones and bacterial DNA binding protein II), which together form chromosomes. The endogenous cellular gene can be in a transcriptionally active or inactive state in the native chromatin environment.

A "developmentally silent gene" or an "inactive gene" refers to a gene whose expression is repressed or not activated, i.e., turned off, in certain cell types, during certain developmental stages of a cell type, or during certain time periods in a cell type. Examples of developmentally inactive genes include EPO, GATA, interleukin family proteins, GM-CSF, MyoD, eutrophin, and fetal hemoglobins gamma and delta.

The phrase "adjacent to a transcription initiation site" refers to a target site that is within about 50 bases either upstream or downstream of a transcription initiation site. "Upstream" of a transcription initiation site refers to a target site that is more than about 50 bases 5' of the transcription initiation site (i.e., in the non-transcribed region of the gene).

The term "zinc finger protein" or "ZFP" refers to a protein having DNA binding domains that are stabilized by zinc. The individual DNA binding domains are typically referred to as "fingers" A zinc finger protein has at least one finger, typically two fingers, three fingers, four fingers, five fingers, or six fingers or more. Each finger binds from two to four base pairs of DNA, typically three or four base pairs of DNA. A zinc finger protein binds to a nucleic acid sequence called a target site or target segment. Each finger typically comprises an approximately 30 amino acid, zinc-coordinating, DNA-binding subdomain. An exemplary motif characterizing one class of these proteins (Cys_2His_2 class) is -Cys-$(X)_{2-4}$-Cys-$(X)_{12}$-His-$(X)_{3-5}$-His (where X is any amino acid). Studies have demonstrated that a single zinc finger of this class consists of an alpha helix containing the two invariant histidine residues co-ordinated with zinc along with the two cysteine residues of a single beta turn (*see, e.g.*, Berg & Shi, *Science* 271:1081-1085 (1996)).

A "target site" is the nucleic acid sequence recognized by a zinc finger protein. A single target site typically has about four to about ten or more base pairs. Typically, a two-fingered zinc finger protein recognizes a four to seven base pair target

six to ten base pair target site, a six
nine to ten base pair target sites, and
rget site is in any position that allows
r downstream of the transcription
criptional regulation element such as
:e elements, nuclear receptor
olymerase pause sites; and
tes" refers to non-overlapping target

it for the compound, i.e., the
ein) that gives half maximal binding
und molecules are bound to the
K_d), as measured using a given assay
ssay system used to measure the K_d
easure of the actual K_d of the ZFP.
accurate measurement of the actual
FPs of the invention is measured

target site is expected to provide specificity in the human genome, as a target site of that size should occur only once in every $3x10^{10}$ base pairs, and the size of the human genome is $3.5x10^9$ base pairs (*see, e.g.*, Liu *et al., PNAS* 94:5525-5530 (1997)). In another embodiment, the ZFPs are non-covalently associated, through a leucine zipper, a STAT protein N-terminal domain, or the FK506 binding protein (*see, e.g.*, O'Shea, *Science* 254: 539 (1991), Barahmand-Pour *et al., Curr. Top. Microbiol. Immunol.* 211:121-128 (1996); Klemm *et al., Annu. Rev. Immunol.* 16:569-592 (1998); Ho *et al., Nature* 382:822-826 (1996)).

In another embodiment, the ZFP is linked to at least one or more regulatory domains, described below. Preferred regulatory domains include transcription factor repressor or activator domains such as KRAB and VP16, co-repressor and co-activator domains, DNA methyl transferases, histone acetyltransferases, histone deacetylases, and endonucleases such as FokI. For repression of gene expression, typically the expression of the gene is reduced by about 20% (i.e., 80% of non-ZFP modulated expression), more preferably by about 50% (i.e., 50% of non-ZFP modulated expression), more preferably by about 75-100% (i.e., 25% to 0% of non-ZFP modulated expression). For activation of gene expression, typically expression is activated by about 1.5 fold (i.e., 150% of non-ZFP modulated expression), preferably 2 fold (i.e., 200% of non-ZFP modulated expression), more preferably 5-10 fold (i.e., 500-1000% of non-ZFP modulated expression), up to at least 100 fold or more.

The expression of engineered ZFP activators and repressors can be also controlled by systems typified by the tet-regulated systems and the RU-486 system (*see, e.g.*, Gossen & Bujard, *PNAS* 89:5547 (1992); Oligino *et al., Gene Ther.* 5:491-496 (1998); Wang *et al., Gene Ther.* 4:432-441 (1997); Neering *et al., Blood* 88:1147-1155 (1996); and Rendahl *et al., Nat. Biotechnol.* 16:757-761 (1998)). These impart small molecule control on the expression of the ZFP activators and repressors and thus impart small molecule control on the target gene(s) of interest. This beneficial feature could be used in cell culture models, in gene therapy, and in transgenic animals and plants.

Definitions

As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

desaturase, delta-9 desaturase, delta-15 desaturase, acetyl-CoA carboxylase, acyl-ACP-thioesterase, ADP-glucose pyrophosphorylase, starch synthase, cellulose synthase, sucrose synthase, senescence-associated genes, heavy metal chelators, fatty acid hydroperoxide lyase, viral genes, protozoal genes, fungal genes, and bacterial genes. In general, suitable genes to be regulated include cytokines, lymphokines, growth factors, mitogenic factors, chemotactic factors, onco-active factors, receptors, potassium channels, G-proteins, signal transduction molecules, and other disease-related genes. Preferred developmentally inactive genes include EPO, GATA, interleukin family proteins, GM-CSF, MyoD, eutrophin, and fetal hemoglobins gamma and delta.

A general theme in transcription factor function is that simple binding and sufficient proximity to the promoter are all that is generally needed. Exact positioning relative to the promoter, orientation, and within limits, distance, do not matter greatly for expression modulation by a ZFP. This feature allows considerable flexibility in choosing sites for constructing artificial transcription factors. The target site recognized by the ZFP therefore can be any suitable site in the target gene that will allow activation or repression of gene expression by a ZFP, optionally linked to a regulatory domain. Preferred target sites include regions adjacent to, downstream, or upstream of the transcription start site. In addition, target sites can also be located in enhancer regions, repressor sites, RNA polymerase pause sites, and specific regulatory sites (e.g., SP-1 sites, hypoxia response elements, nuclear receptor recognition elements, p53 binding sites), sites in the cDNA encoding region or in an expressed sequence tag (EST) coding region. As described below, typically each finger recognizes 2-4 base pairs, with a two finger ZFP binding to a 4 to 7 bp target site, a three finger ZFP binding to a 6 to 10 base pair site, and a six finger ZFP binding to two adjacent target sites, each target site having from 6-10 base pairs.

As described herein, two ZFPs can be administered to a cell, recognizing either the same target endogenous cellular gene, or different target endogenous cellular gene. The first ZFP optionally is associated with the second ZFP, either covalently or non-covalently. Recognition of adjacent target sites by either associated or individual ZFPs can be used to produce cooperative binding of the ZFPs, resulting in an affinity that is greater than the affinity of the ZFPs when individually bound to their target site.

In one embodiment, two ZFPs are produced as a fusion protein linked by an amino acid linker, and the resulting six finger ZFP recognizes an approximately 18 base pair target site (*see, e.g.*, Liu *et al.*, *PNAS* 94:5525-5530 (1997)). An 18 base pair

chromatin environment. The present invention thus provides zinc finger DNA binding proteins that have been engineered to specifically recognize, with high efficacy, endogenous cellular genes. The experiments described herein demonstrate that a 3 finger ZFP with a target site affinity of less than about 10 nM (VEGF1) can be used to effectively activate or repress activity of an endogenous gene. Furthermore, a 6 finger ZFP (VEGF3a/1) was also shown to effectively repress activity of an endogenous gene. Finally, three finger ZFP can be used to activate endogenous EPO, a developmentally inactive gene. Preferably, the ZFPs of the invention exhibit high affinity for their target sites, with K_ds of less than about 100 nM, preferably less than about 50 nM, most preferably less than about 25 nM or lower.

As a result, the ZFPs of the invention can be used to regulate endogenous gene expression, both through activation and repression of endogenous gene transcription. The ZFPs can also be linked to regulatory domains, creating chimeric transcription factors to activate or repress transcription. In one preferred embodiment, the methods of regulation use ZFPs with a K_d of less than about 25 nM to activate or repress gene transcription. The ZFPs of the invention therefore can be used to repress transcription of an endogenous cellular gene by 20% or more, and can be used to activate transcription of an endogenous cellular gene by about 1.5 fold or more.

Such methods of regulating gene expression allow for novel human and mammalian therapeutic applications, e.g., treatment of genetic diseases, cancer, fungal, protozoal, bacterial, and viral infection, ischemia, vascular disease, arthritis, immunological disorders, etc., as well as providing means for functional genomics assays, and means for developing plants with altered phenotypes, including disease resistance, fruit ripening, sugar and oil composition, yield, and color.

As described herein, ZFPs can be designed to recognize any suitable target site, for regulation of expression of any endogenous gene of choice. Examples of endogenous genes suitable for regulation include VEGF, CCR5, ERα, Her2/Neu, Tat, Rev, HBV C, S, X, and P, LDL-R, PEPCK, CYP7, Fibrinogen, ApoB, Apo E, Apo(a), renin, NF-κB, I-κB, TNF-α, FAS ligand, amyloid precursor protein, atrial naturetic factor, ob-leptin, ucp-1, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-12, G-CSF, GM-CSF, Epo, PDGF, PAF, p53, Rb, fetal hemoglobin, dystrophin, eutrophin, GDNF, NGF, IGF-1, VEGF receptors flt and flk, topoisomerase, tel merase, bcl-2, cyclins, angiostatin, IGF, ICAM-1, STATS, c-myc, c-myb, TH, PTI-1, polygalacturonase, EPSP synthase, FAD2-1, delta-12

Figure 7. Co-transfection data showing activation of luciferase reporter activity via VEGF-VP16 protein expression. Error bars show the standard deviation of triplicate transfections. pGL3-P (reporter with no VEGF target); pcDNA (empty effector vector control); pVFR3-4x (VEGF reporter plasmid); VEGF1 (VEGF1-VP16); VEGF3a (VEGF3a-VP16); VEGF3a/1 (VEGF3a/1-VP16).

Figure 8. VEGF ELISA data showing repression of endogenous VEGF gene expression due to transfection of a VEGF ZFP-KRAB effector plasmid. DFX treated (control nontransfected Dfx treated cells; No ZFP (pcDNA-control), VEGF 1 (VEGF1-KRAB), VEGF 3a/1 (VEGF3a/1-KRAB), CCR5 (CCR5-KRAB); Mock uninduced (mock transfected cells untreated with DFX). Error bars show the standard deviation of duplicate transfections.

Figure 9. VEGF ELISA data showing activation of endogenous VEGF gene expression due to transfection of a VEGF ZFP-VP16 effector plasmid. Mock (mock transfected cells); No ZFP (NVF-control), VEGF 1 (VEGF1-VP16), VEGF 3a/1 (VEGF3a/1-VP16). Error bars show the standard deviation of duplicate transfections.

Figure 10. RNase protection assay showing changes in VEGF specific mRNA by VEGF-specific ZFPs. Panel A: Activation of VEGF mRNA, NVF-Control (no ZFP), VEGF1-NVF (VEGF1-VP16), CCR5-5-NVF (CCR5-VP16), CCR5-3-NVF (CCR5-VP16). Panel B: Repression of VEGF mRNA. NKF-Control (no ZFP), VEGF1-NKF (VEGF1-KRAB), VEGF3a/1-NKF (VEGF3a/1-KRAB), CCR5-3-NKF (CCR5-KRAB). The size of the 148 nucleotide VEGF specific band is indicated by an arrow. The VEGF specific probe was synthesized from a human angiogenesis multi-probe template set (Pharmingen).As a control, signals from the housekeeping genes L32 and GAPDH are shown (arrows).

Figure 11a-b: Figure 11a shows activation of endogenous EPO gene expression by measuring EPO production in Hep3B and 293 cells, as measuring using ELISA. Figure 11b shows activation of endogenous EPO gene expression in Hep3B cells and 293 cells by measuring mRNA expression.

DETAILED DESCRIPTION OF THE INVENTION

Introduction

The present application demonstrates for the first time that ZFPs can be used to regulate expression of an endogenous cellular gene that is present in its native

target site is adjacent to an RNA polymerase pause site downstream of a transcription initiation site of the endogenous cellular gene.

In another embodiment, the zinc finger protein comprises an SP-1 backbone. In one embodiment, the zinc finger protein comprises a regulatory domain and is humanized.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: PCR amplification scheme for production of ZFP-encoding synthetic genes.

Figure 2. Expression and purification of typical ZFPs. Fig. 2A: Unfused ZFP before induction (lane 1), after induction (lane 2), and after purification (lane 3). Fig. 2B: MBP-VEGF expression before induction (lane 1), after induction (lane 2), and after French Press lysis (lane 3). Fig. 2C: Purification of MBP-VEGF by amylose affinity column showing flow-through (FT), and initial fractions (1-4). Fraction 2 was used for electrophoretic mobility shift assays ("EMSA"). M, molecular weight markers.

Figure 3. Typical EMSA experiment with MBP fused ZFP. MBP-VEGF1 protein was bound to labeled duplex DNA as described in the text. A three-fold protein dilution series was carried out; each point represents the percent shifted at that particular protein concentration plotted on a semi-log graph. Quantitation was by phosphorimager. In this case, the protein concentration yielding 50% of maximum shift (the apparent K_d) was 2 nM.

Figure 4. Off-rate experiment comparing VEGF1 to VEGF3a/1. Protein-DNA complexes were pre-formed and incubated with a 1000-fold excess of unlabeled oligonucleotide. Samples were electrophoresed at various times and the amount of shifted product was measured by phosphorimager. Curve fitting was used to calculate the indicated complex half-lives.

Figure 5. Typical expression vector used for transient ZFP expression in mammalian cells.

Figure 6. Co-transfection data showing repression of luciferase reporter activity via VEGF-KRAB protein expression. Error bars show the standard deviation of triplicate transfections. pGL3-C (reporter vector control); pVFR1-4x (VEGF reporter plasmid); VEGF1 (VEGF1-KRAB); VEGF3a (VEGF3a-KRAB); VEGF3a/1 (VEGF3a/1-KRAB).

In one embodiment, the endogenous cellular gene is a selected from the group consisting of VEGF, ERα, IGF-I, c-myc, c-myb, ICAM, Her2/Neu, FAD2-1, EPO, GM-CSF, GDNF, and LDL-R. In another embodiment, the endogenous cellular gene is a developmentally silent or otherwise inactive gene, e.g., EPO, GATA, interleukin family proteins, GM-CSF, MyoD, eutrophin, and fetal hemoglobins gamma and delta. In another embodiment, the regulatory domain is selected from the group consisting of a transcriptional repressor, a transcriptional activator, an endonuclease, a methyl transferase, a histone acetyltransferase, and a histone deacetylase.

In one embodiment, the cell is selected from the group consisting of animal cell, a plant cell, a bacterial cell, a protozoal cell, or a fungal cell. In another embodiment, the cell is a mammalian cell. In another embodiment, the cell is a human cell.

In one embodiment, the method further comprises the step of first administering to the cell a delivery vehicle comprising the zinc finger protein, wherein the delivery vehicle comprises a liposome or a membrane translocation polypeptide.

In one embodiment, the zinc finger protein is encoded by a zinc finger protein nucleic acid operably linked to a promoter, and the method further comprises the step of first administering the nucleic acid to the cell in a lipid:nucleic acid complex or as naked nucleic acid. In another embodiment, the zinc finger protein is encoded by an expression vector comprising a zinc finger protein nucleic acid operably linked to a promoter, and the method further comprises the step of first administering the expression vector to the cell. In another embodiment, the expression vector is a viral expression vector. In another embodiment, the expression vector is a retroviral expression vector, an adenoviral expression vector, a DNA plasmid expression vector, or an AAV expression vector.

In one the zinc finger protein is encoded by a nucleic acid operably linked to an inducible promoter. In another embodiment, the zinc finger protein is encoded by a nucleic acid operably linked to a weak promoter.

In one embodiment, the cell comprises less than about $1.5x10^6$ copies of the zinc finger protein.

In one embodiment, the target site is upstream of a transcription initiation site of the endogenous cellular gene. In another embodiment, the target site is adjacent to a transcription initiation site of the endogenous cellular gene. In another embodiment, the

the K_d of the zinc finger protein is less than about 25 nM; thereby activating expression of the endogenous cellular gene to at least about 150%.

In another aspect, the present invention provides a method of activating expression of an endogenous cellular gene, the method comprising the step of: contacting a target site in the endogenous cellular gene with a fusion zinc finger protein comprising six fingers and a regulatory domain, wherein the K_d of the zinc finger protein is less than about 25 nM; thereby activating expression of the endogenous cellular gene to at least about 150%.

In one embodiment, expression of the endogenous cellular gene is activated to at least about 200-500%. In another embodiment, activation of gene expression prevents repression of gene expression.

In another aspect, the present invention provides a method of modulating expression of an endogenous cellular gene in a cell, the method comprising the step of: contacting a first target site in the endogenous cellular gene with a first zinc finger protein; thereby modulating expression of the endogenous cellular gene.

In one embodiment, the zinc finger protein has two, three, four, five, or six fingers.

In another aspect, the present invention provides a method of modulating expression of an endogenous cellular gene in a cell, the method comprising the step of: contacting a target site in the endogenous cellular gene with a fusion zinc finger protein comprising six fingers and a regulatory domain; thereby modulating expression of the endogenous cellular gene.

In one embodiment, the step of contacting further comprises contacting a second target site in the endogenous cellular gene with a second zinc finger protein. In another embodiment, the first and second target sites are adjacent. In another embodiment, the first and second zinc finger proteins are covalently linked. In another embodiment, the first zinc finger protein is a fusion protein comprising a regulatory domain. In another embodiment, the first zinc finger protein is a fusion protein comprising at least two regulatory domains. In another embodiment, the first and second zinc finger proteins are fusion proteins, each comprising a regulatory domain. In another embodiment, the first and second zinc finger protein are fusion proteins, each comprising at least two regulatory domains.

is randomly integrated into the genome and is not found in a native chromatin environment as compared to an endogenous gene. In contrast, specific regulation of an endogenous cellular gene in its native chromatin environment using a ZFP has not yet been demonstrated in the art.

SUMMARY OF THE INVENTION

The present invention thus provides for the first time methods of regulating endogenous cellular gene expression, where the endogenous genes are in their native chromatin environment, in contrast to genes that have been transiently expressed in a cell, or those that have been exogenously integrated into the genome. In addition, the present invention provides for the first time activation of a developmentally silent, endogenous gene. In one preferred embodiment, the methods of regulation use ZFPs with a K_d of less than about 25 nM to activate or repress gene transcription. The ZFPs of the invention therefore can be used to repress transcription of an endogenous cellular gene by 20% or more, and can be used to activate transcription of an endogenous cellular gene by about 1.5 fold or more.

In one aspect, the present invention provides a method of inhibiting expression of an endogenous cellular gene in a cell, the method comprising the step of: contacting a first target site in the endogenous cellular gene with a first zinc finger protein, wherein the K_d of the zinc finger protein is less than about 25 nM; thereby inhibiting expression of the endogenous cellular gene by at least about 20%.

In another aspect, the present invention provides a method of inhibiting expression of an endogenous cellular gene in a cell, the method comprising the step of: contacting a target site in the endogenous cellular gene with a fusion zinc finger protein comprising six fingers and a regulatory domain, wherein the K_d of the zinc finger protein is less than about 25 nM; thereby inhibiting expression of the endogenous cellular gene by at least about 20%.

In one embodiment, expression of the endogenous cellular gene is inhibited by at least about 75%-100%. In another embodiment, the inhibition of gene expression prevents gene activation.

In another aspect, the present invention provides a method of activating expression of an endogenous cellular gene, the method comprising the step of: contacting a first target site in the endogenous cellular gene with a first zinc finger protein, wherein

its two components. The authors then constructed a reporter vector containing a luciferase gene operably linked to a promoter and a hybrid site for the chimeric DNA binding protein in proximity to the promoter. The authors reported that their chimeric DNA binding protein could activate or repress expression of the luciferase gene.

Liu *et al.*, *PNAS* 94:5525-5530 (1997) report forming a composite ZFP by using a peptide spacer to link two component ZFPs, each having three fingers. The composite protein was then further linked to transcriptional activation or repression domains. It was reported that the resulting chimeric protein bound to a target site formed from the target segments bound by the two component ZFPs. It was further reported that the chimeric ZFP could activate or repress transcription of a reporter gene when its target site was inserted into a reporter plasmid in proximity to a promoter operably linked to the reporter.

Beerli *et al.*, *PNAS* 95:14628-14633 (1998) report construction of a chimeric six finger ZFP fused to either a KRAB, ERD, or SID transcriptional repressor domain, or the VP16 or VP64 transcriptional activation domain. This chimeric ZFP was designed to recognize an 18 bp target site in the 5' untranslated region of the human erbB-2 gene. Using this construct, the authors of this study report both activation and repression of a transiently expressed reporter luciferase construct linked to the erbB-2 promoter.

In addition, a recombinant ZFP was reported to repress expression of an integrated plasmid construct encoding a *bcr-abl* oncogene (Choo *et al.*, *Nature* 372:642-645 (1994)). The target segment to which the ZFPs bound was a nine base sequence GCA GAA GCC chosen to overlap the junction created by a specific oncogenic translocation fusing the genes encoding bcr and abl. The intention was that a ZFP specific to this target site would bind to the oncogene without binding to abl or bcr component genes. The authors used phage display to select a variant ZFP that bound to this target segment. The variant ZFP thus isolated was then reported to repress expression of a stably transfected *bcr-abl* construct in a cell line.

To date, these methods have focused on regulation of either transiently expressed genes, or on regulation of exogenous genes that have been integrated into the genome. The transiently expressed genes described by Pomerantz *et al.*, Liu *et al.*, and Beerli *et al.* are episomal and are not packaged into chromatin in the same manner as chromosomal genes. Moreover, even the stably expressed gene described by Choo *et al.*

shows that each finger can be superimposed on the next by a periodic rotation and translation of the finger along the main DNA axis. The structure suggests that each finger interacts independently with DNA over 3 base-pair intervals, with side-chains at positions -1, 2 , 3 and 6 on each recognition helix making contacts with respective DNA triplet sub-site. The amino terminus of Zif268 is situated at the 3' end of its DNA recognition subsite. Some zinc fingers can bind to a fourth base in a target segment. The fourth base is on the opposite strand from the other three bases recognized by zinc finger and complementary to the base immediately 3' of the three base subsite.

The structure of the Zif268-DNA complex also suggested that the DNA sequence specificity of a ZFP might be altered by making amino acid substitutions at the four helix positions (-1, 2, 3 and 6) on a zinc finger recognition helix. Phage display experiments using zinc finger combinatorial libraries to test this observation were published in a series of papers in 1994 (Rebar *et al.*, *Science* 263:671-673 (1994); Jamieson *et al.*, *Biochemistry* 33:5689-5695 (1994); Choo *et al.*, *PNAS* 91:11163-11167 (1994)). Combinatorial libraries were constructed with randomized side-chains in either the first or middle finger of Zif268 and then isolated with an altered Zif268 binding site in which the appropriate DNA sub-site was replaced by an altered DNA triplet. Correlation between the nature of introduced mutations and the resulting alteration in binding specificity gave rise to a partial set of substitution rules for rational design of ZFPs with altered binding specificity.

Greisman & Pabo, *Science* 275:657-661 (1997) discuss an elaboration of a phage display method in which each finger of a zinc finger protein is successively subjected to randomization and selection. This paper reported selection of ZFPs for a nuclear hormone response element, a p53 target site and a TATA box sequence.

Recombinant ZFPs have been reported to have the ability to regulate gene expression of transiently expressed reporter genes in cultured cells (*see, e.g.*, Pomerantz *et al., Science* 267:93-96 (1995); Liu *et al., PNAS* 94:5525-5530 1997); and Beerli *et al., PNAS* 95:14628-14633 (1998)).

For example, Pomerantz *et al., Science* 267:93-96 (1995) report an attempt to design a novel DNA binding protein by fusing two fingers from Zif268 with a homeodomain from Oct-1. The hybrid protein was then fused with either a transcriptional activator or repressor domain for expression as a chimeric protein. The chimeric protein was reported to bind a target site representing a hybrid of the subsites of

Gene expression is normally controlled through alterations in the function of sequence specific DNA binding proteins called transcription factors. These bind in the general proximity (although occasionally at great distances) of the point of transcription initiation of a gene. They act to influence the efficiency of formation or function of a transcription initiation complex at the promoter. Transcription factors can act in a positive fashion (transactivation) or in a negative fashion (transrepression).

Transcription factor function can be constitutive (always "on") or conditional. Conditional function can be imparted on a transcription factor by a variety of means, but the majority of these regulatory mechanisms depend of the sequestering of the factor in the cytoplasm and the inducible release and subsequent nuclear translocation, DNA binding and transactivation (or repression). Examples of transcription factors that function this way include progesterone receptors, sterol response element binding proteins (SREBPs) and NF-kappa B. There are examples of transcription factors that respond to phosphorylation or small molecule ligands by altering their ability to bind their cognate DNA recognition sequence (Hou *et al.*, *Science* 256:1701 (1994); Gossen & Bujard, *PNAS* 89:5547 (1992); Oligino *et al.*, *Gene Ther.* 5:491-496 (1998); Wang *et al.*, *Gene Ther.* 4:432-441 (1997); Neering *et al.*, *Blood* 88:1147-1155 (1996); and Rendahl *et al.*, *Nat. Biotechnol.* 16:757-761 (1998)). This mechanism is common in prokaryotes but somewhat less common in eukaryotes.

Zinc finger proteins ("ZFPs") are proteins that can bind to DNA in a sequence-specific manner. Zinc fingers were first identified in the transcription factor TFIIIA from the oocytes of the African clawed toad, *Xenopus laevis*. ZFPs are widespread in eukaryotic cells. An exemplary motif characterizing one class of these proteins (C_2H_2 class) is -Cys-$(X)_{2-4}$-Cys-$(X)_{12}$-His-$(X)_{3-5}$-His (where X is any amino acid). A single finger domain is about 30 amino acids in length and several structural studies have demonstrated that it contains an alpha helix containing the two invariant histidine residues co-ordinated through zinc with the two cysteines of a single beta turn. To date, over 10,000 zinc finger sequences have been identified in several thousand known or putative transcription factors. ZFPs are involved not only in DNA-recognition, but also in RNA binding and protein-protein binding. Current estimates are that this class of molecules will constitute about 2% of all human genes.

The X-ray crystal structure of Zif268, a three-finger domain from a murine transcription factor, has been solved in complex with its cognate DNA-sequence and

25:3559(1997); Wolffe *et al., PNAS 96:5894(1999)*). Examples of the therapeutic benefit for expression of such genes include activation of developmentally silent fetal hemoglobin genes to treat sickle cell disease and the activation of eutrophin to treat muscular dystrophy. In addition, pathogenic organisms such as viruses, bacteria, fungi, and protozoa could be controlled by altering gene expression. There is thus a clear unmet need for therapeutic approaches that act through sequence-specific regulation of disease-related genes.

In addition to the direct therapeutic utility provided by the ability to manipulate gene expression, this ability can be used experimentally to determine the function of a gene of interest. One common existing method for experimentally determining the function of a newly discovered gene is to clone its cDNA into an expression vector driven by a strong promoter and measure the physiological consequence of its over-expression in a transfected cell. This method is labor intensive and does not address the physiological consequences of down-regulation of a target gene. Simple methods allowing the selective over and under-expression of uncharacterized genes would be of great utility to the scientific community. Methods that permit the regulation of genes in cell model systems, transgenic animals and transgenic plants would find widespread use in academic laboratories, pharmaceutical companies, genomics companies and in the biotechnology industry.

An additional use of tools permitting the manipulation of gene expression is in the production of commercially useful biological products. Cell lines, transgenic animals and transgenic plants could be engineered to over-express a useful protein product. The production of erythropoietin by such an engineered cell line serves as an example. Likewise, production from metabolic pathways might be altered or improved by the selective up or down-regulation of a gene encoding a crucial enzyme. An example of this is the production of plants with altered levels of fatty acid saturation.

Methods currently exist in the art, which allow one to alter the expression of a given gene, e.g., using ribozymes, antisense technology, small molecule regulators, over-expression of cDNA clones, and gene-knockouts. These methods have to date proven to be generally insufficient for many applications and typically have not demonstrated either high target efficacy or high specificity *in vivo*. For useful experimental results and therapeutic treatments, these characteristics are desired.

REGULATION OF ENDOGENOUS GENE EXPRESSION IN CELLS USING ZINC FINGER PROTEINS

CROSS-REFERENCES TO RELATED APPLICATIONS

This application is a continuation-in-part of and claims the benefit of USSN 09/229,037, filed January 12, 1999. This application is also related to USSN 09/229,007, filed January 12, 1999, herein both incorporated by reference in their entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This invention was made with government support under Grant No. 1 R43 DK52251-01, awarded by the National Institutes of Health. The government has certain rights in this invention.

FIELD OF THE INVENTION

The present invention provides methods for regulating gene expression of endogenous genes using recombinant zinc finger proteins.

BACKGROUND OF THE INVENTION

Many pathophysiological processes are the result of aberrant gene expression. Examples include the inappropriate activation of proinflammatory cytokines in rheumatoid arthritis, under-expression of the hepatic LDL receptor in hypercholesteremia, over-expression of proangiogenic factors, and under-expression of antiangiogenic factors in solid tumor growth. If therapeutic methods for control of gene expression existed, many of these pathologies could be more optimally treated. In another example, developmentally silent or otherwise inactive genes could be activated in order to treat a particular disease state. Inactive genes are repressed via several mechanisms, including chromatin structure, specific *cis*-acting repressors, and DNA methylation (Travers, *Cell* 96:311(1996); Beato & Eisfeld, *Nucleic Acids Res.*

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 July 2000 (20.07.2000)

PCT

(10) International Publication Number
WO 00/41566 A1

(51) International Patent Classification[7]: A01N 63/00

(21) International Application Number: PCT/US00/00409

(22) International Filing Date: 6 January 2000 (06.01.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/229,037 12 January 1999 (12.01.1999) US

(71) Applicant *(for all designated States except US)*: SANAGAMO BIOSCIENCES, INC. [US/US]; Suite A100, 501 Canal Boulevard, Richmond, CA 94804 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **COX, George, Norbert, III** [US/US]; 678 West Willow Street, Louisville, CO 80027 (US). **CASE, Casey, Christopher** [US/US]; 605 Barroilhet Avenue, San Mateo, CA 94402 (US). **EISENBERG, Stephen, P.** [US/US]; 5664 Pennsylvania Place, Boulder, CO 80303 (US). **JARVIS, Eric, Edward** [US/US]; 3720 Smuggler Place, Boulder, CO 80303 (US). **SPRATT, Sharon, Kaye** [US/US]; 107 Edmondale Court, Vacaville, CA 95688 (US).

(74) Agents: **PARENT, Annette, S.** et al.; Townsend & Townsend & Crew LLP, 8th Floor, Two Embarcadero Center, San Francisco, CA 94111-3834 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(48) Date of publication of this corrected version:
26 July 2001

(15) Information about Correction:
see PCT Gazette No. 30/2001 of 26 July 2001, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/41566 A1

(54) Title: REGULATION OF ENDOGENOUS GENE EXPRESSION IN CELLS USING ZINC FINGER PROTEINS

(57) Abstract: The present invention provides methods for modulating expression of endogenous cellular genes using recombinant zinc finger proteins.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/00409

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : A01N 63/00
US CL : 424/93.2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 424/93.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
CAPLUS, MEDLINE, SCISEARCH, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BEERLI et al. Toward controlling gene expression at will: Specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc. Natl. Acad. Sci. USA. December 1998, Vol.95, pages 14628-14633, see entire document.	1-90

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 March 2000 (24.03.2000)	04 APR 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Jeffrey S. Lundgren Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

Figure 11a. Activation of Endogenous EPO genes by ZFP

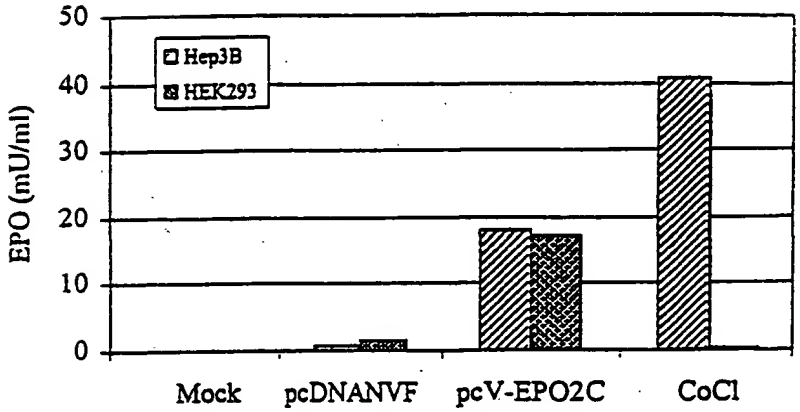


Figure 11b. Effects of ZFP on EPO mRNA expression in Hep3B cells

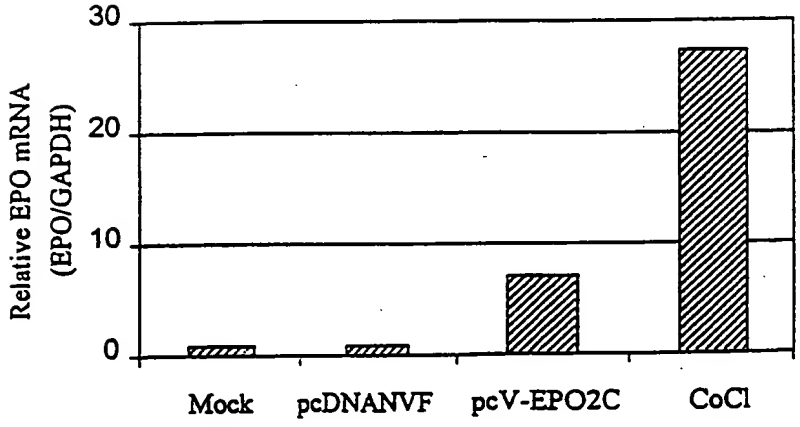


Figure 11

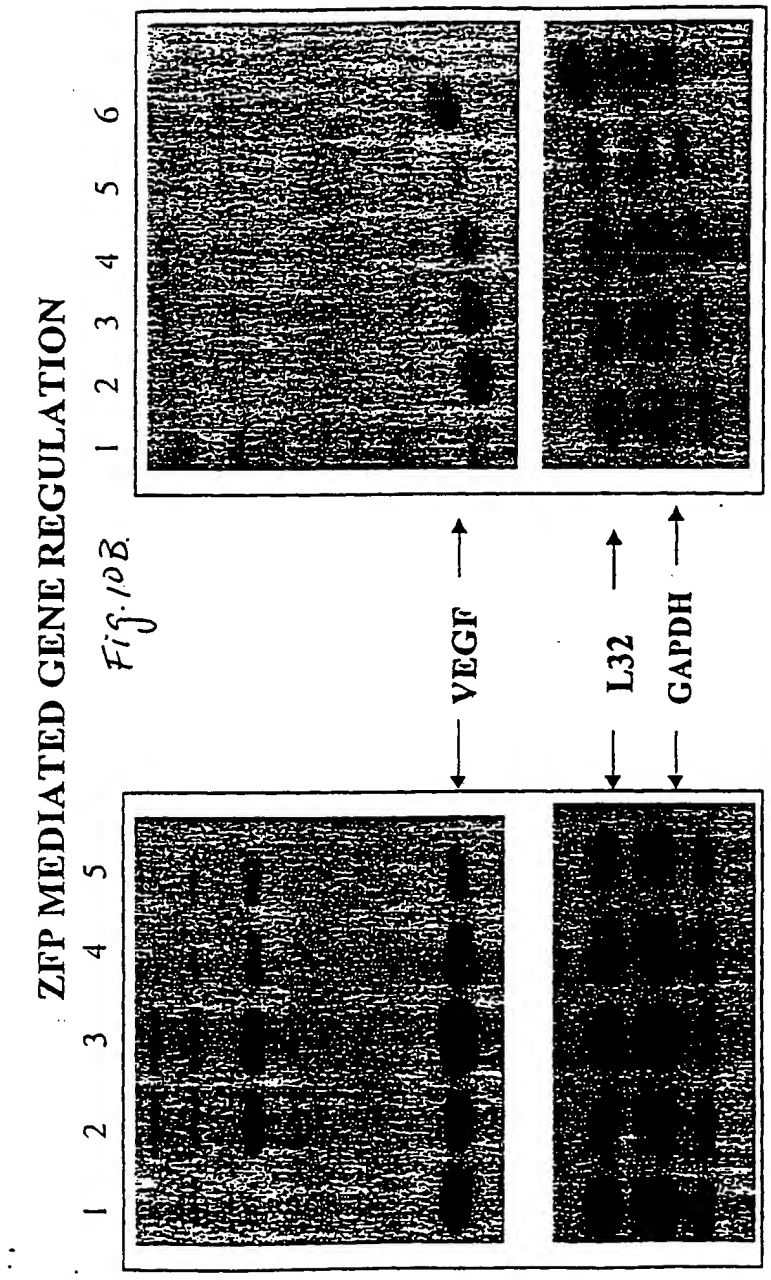


Fig. 10A

Fig. 10B

Figure 10

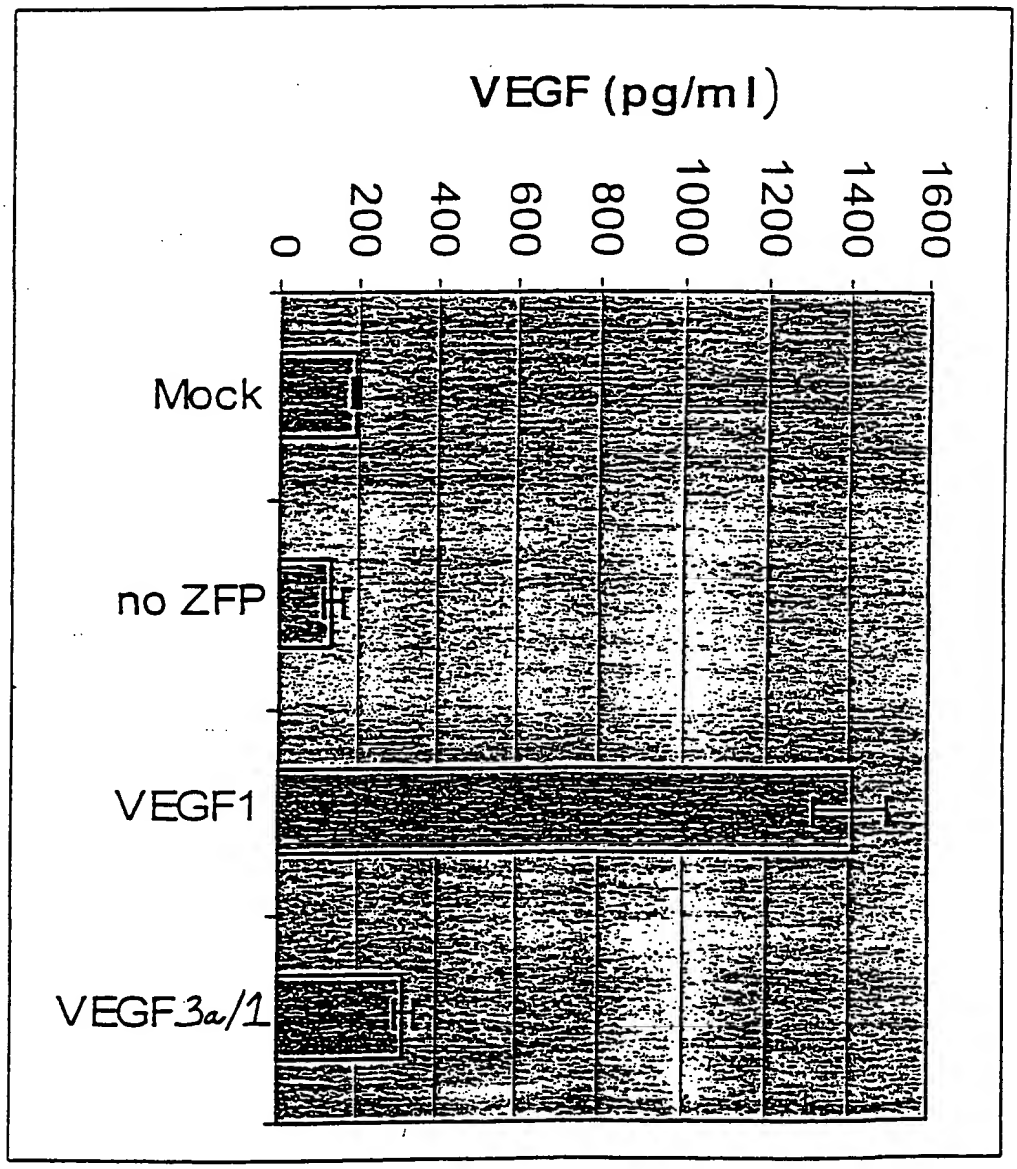

VEGF (pg/ml)
0
200
400
600
800
1000
1200
1400
1600
Mock
no ZFP
VEGF1
VEGF3a/1

Figure 9

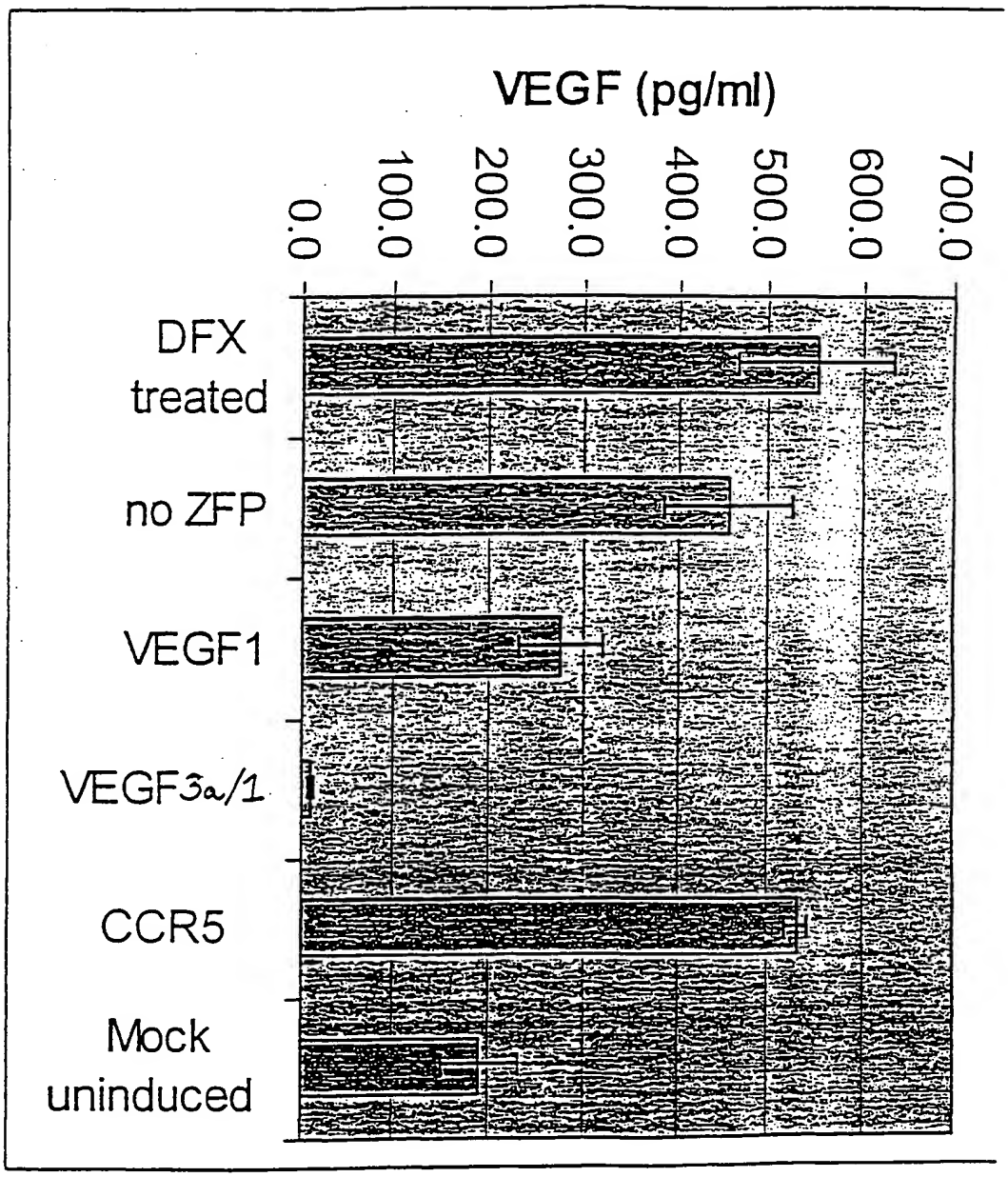

VEGF (pg/ml)
0.0
100.0
200.0
300.0
400.0
500.0
600.0
700.0
DFX treated
no ZFP
VEGF1
VEGF3a/1
CCR5
Mock uninduced

Figure 8

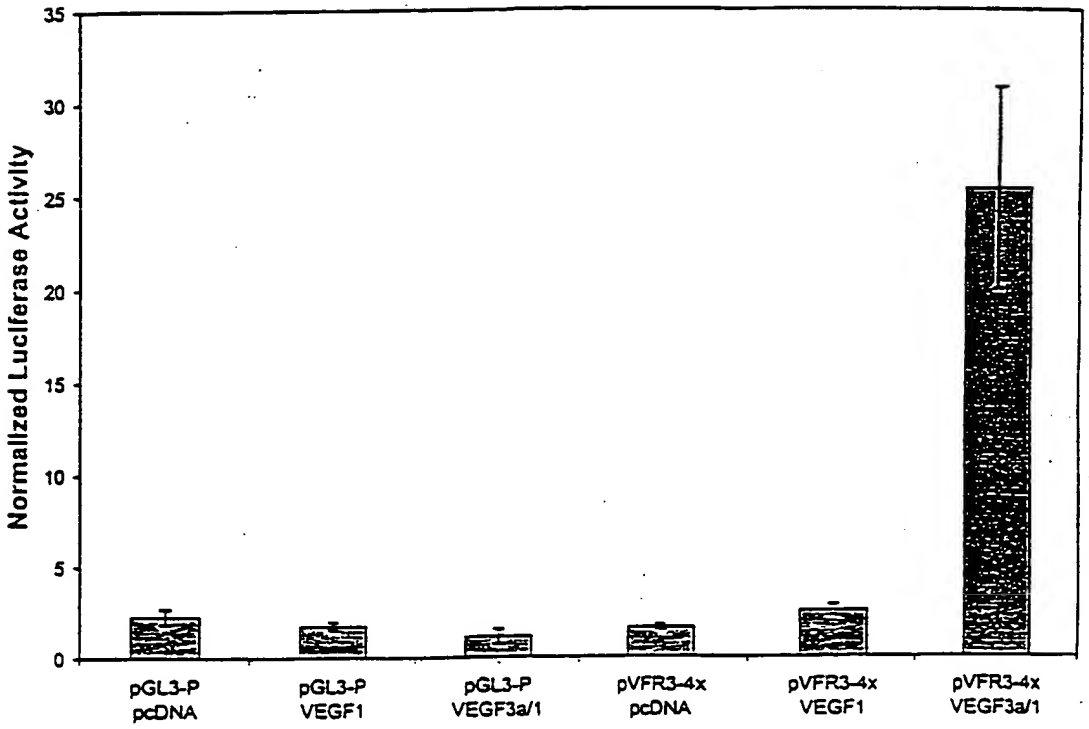

Normallzed Luclferase Activlty
35
30
25
20
15
10
5
0
pGL3-P pcDNA
pGL3-P VEGF1
pGL3-P VEGF3a/1
pVFR3-4x pcDNA
pVFR3-4x VEGF1
pVFR3-4x VEGF3a/1

Figure 7

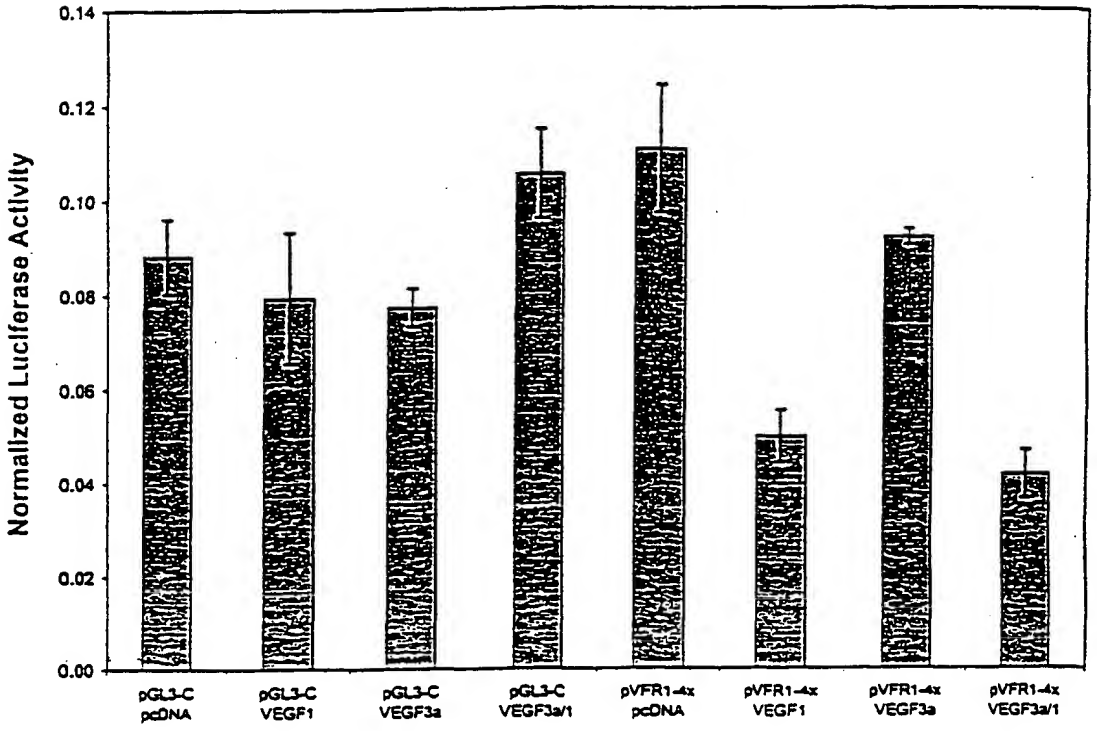

Normalized Luciferase Activity
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
pGL3-C pcDNA
pGL3-C VEGF1
pGL3-C VEGF3a
pGL3-C VEGF3a/1
pVFR1-4x pcDNA
pVFR1-4x VEGF1
pVFR1-4x VEGF3a
pVFR1-4x VEGF3a/1

Figure 6

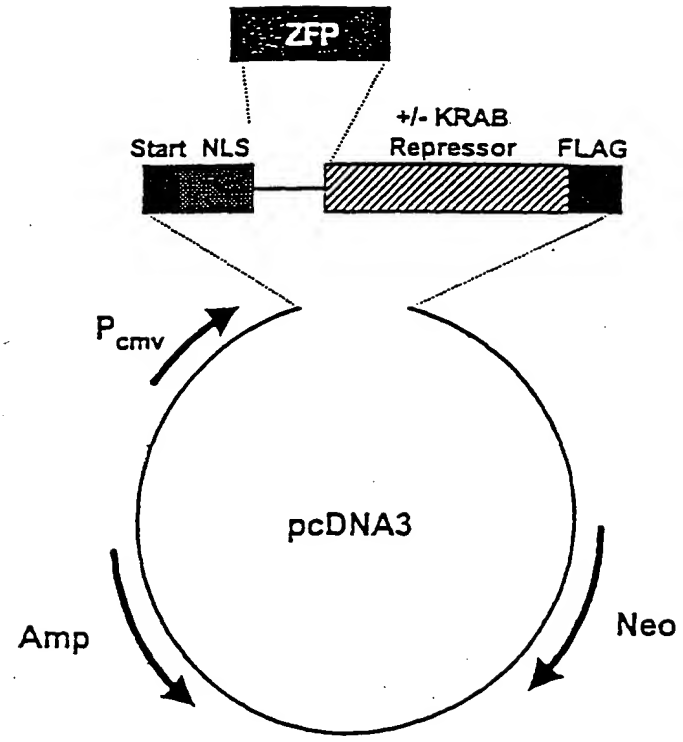


Figure 5

Fig. 4A

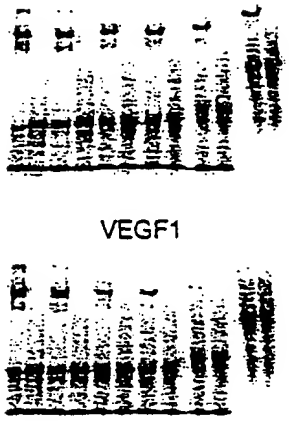


Fig. 4B

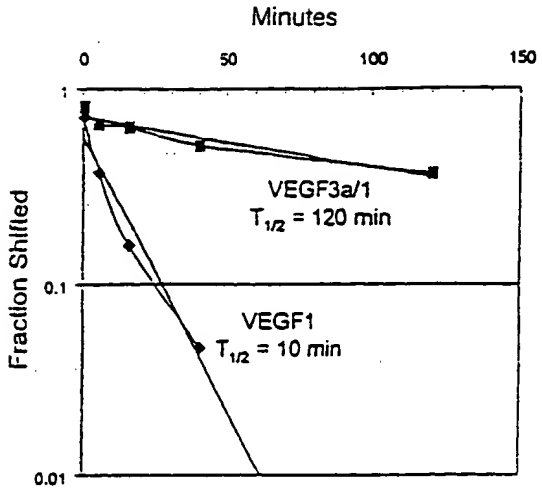


Figure 4

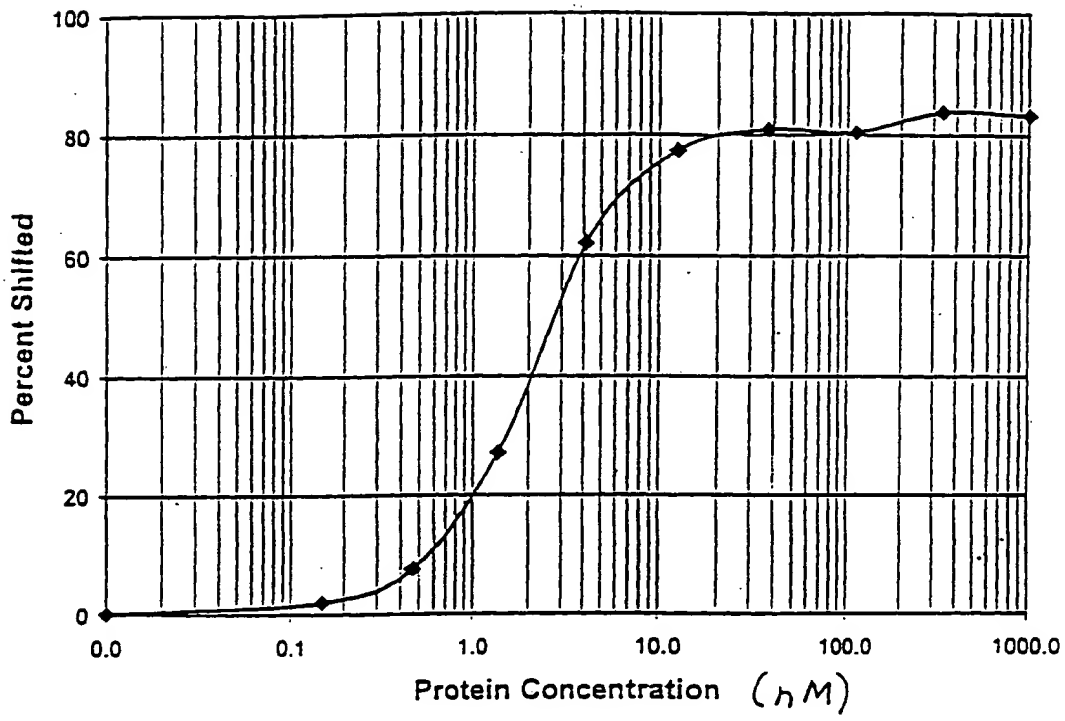

100
80
60
40
20
0
Percent Shifted
0.0
0.1
1.0
10.0
100.0
1000.0
Protein Concentration (nM)

Figure 3

Fig. 2 A

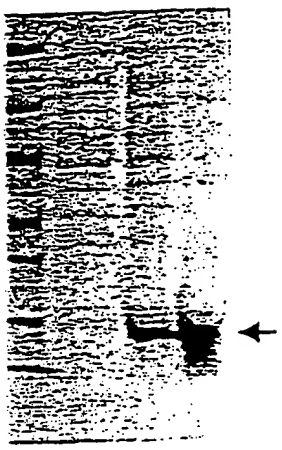


Fig. 2 B

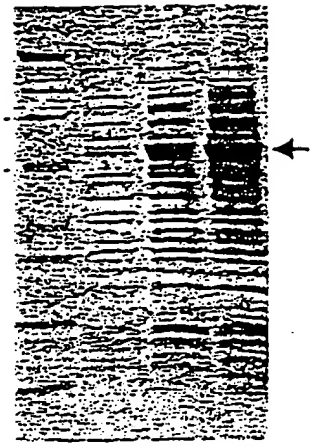


Fig. 2 C

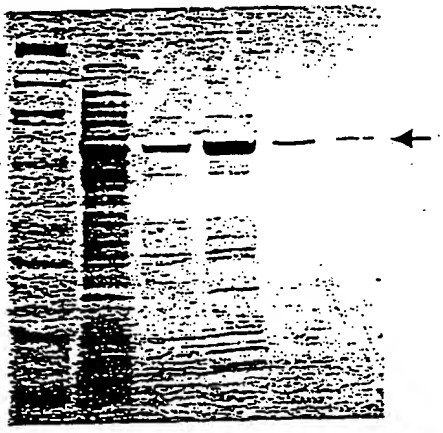


Figure 2

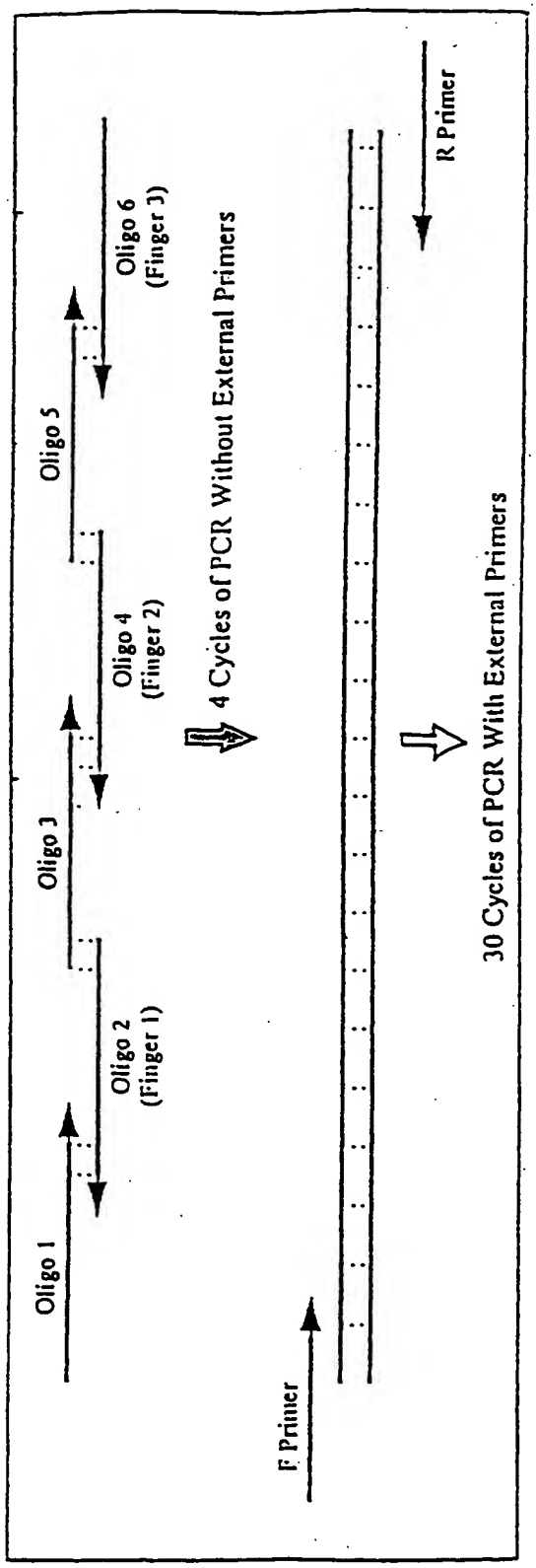

Oligo 1
Oligo 2
(Finger 1)
Oligo 3
Oligo 4
(Finger 2)
Oligo 5
Oligo 6
(Finger 3)
4 Cycles of PCR Without External Primers
F Primer
R Primer
30 Cycles of PCR With External Primers

Figure 1

82. The method of claim 79, wherein the zinc finger protein is encoded by a nucleic acid operably linked to an inducible promoter.

83. The method of claim 79, wherein the zinc finger protein is encoded by a nucleic acid operably linked to a weak promoter.

84. The method of claim 63, wherein the cell comprises less than about $1.5x10^6$ copies of the zinc finger protein.

85. The method of claim 63, wherein the target site is upstream of a transcription initiation site of the endogenous cellular gene.

86. The method of claim 63, wherein the target site is adjacent to a transcription initiation site of the endogenous cellular gene.

87. The method of claim 63, wherein the target site is adjacent to an RNA polymerase pause site downstream of a transcription initiation site of the endogenous cellular gene.

88. The method of claim 63, wherein the zinc finger protein comprises an SP-1 backbone.

89. The method of claim 89, wherein the zinc finger protein comprises a regulatory domain and is humanized.

90. A method of modulating expression of an endogenous cellular gene in a cell, the method comprising the step of:

contacting a target site in the endogenous cellular gene with a fusion zinc finger protein comprising six fingers and a regulatory domain;

thereby modulating expression of the endogenous cellular gene.

72. The method of claim 71, wherein the cell is a mammalian cell

73. The method of claim 72, wherein the cell is a human cell.

74. The method of claim 63, wherein the endogenous cellular gene is a selected from the group consisting of VEGF, ERα, IGF-I, c-myc, c-myb, ICAM, Her2/Neu, FAD2-1, EPO, GM-CSF, GDNF, and LDL-R.

75. The method of claim 63, wherein the endogenous cellular gene is VEGF.

76. The method of claim 67 or 69, wherein the regulatory domain is selected from the group consisting of a transcriptional repressor, a transcriptional activator, an endonuclease, a methyl transferase, a histone acetyltransferase, and a histone deacetylase.

77. The method of claim 63, wherein the method further comprises the step of first administering to the cell a delivery vehicle comprising the zinc finger protein, wherein the delivery vehicle comprises a liposome or a membrane translocation polypeptide.

78. The method of claim 63, wherein the zinc finger protein is encoded by a zinc finger protein nucleic acid operably linked to a promoter, and wherein the method further comprises the step of first administering the nucleic acid to the cell in a lipid:nucleic acid complex or as naked nucleic acid.

79. The method of claim 63, wherein the zinc finger protein is encoded by an expression vector comprising a zinc finger protein nucleic acid operably linked to a promoter, and wherein the method further comprises the step of first administering the expression vector to the cell.

80. The method of claim 79, wherein the expression vector is a viral expression vector.

81. The meth d of claim 79, wherein the expression vector is a retroviral expression vector, an adenoviral expression vector, or an AAV expression vector.

contacting a target site in the endogenous cellular gene with a fusion zinc finger protein comprising six fingers and a regulatory domain, wherein the K_d of the zinc finger protein is less than about 25 nM;

thereby activating expression of the endogenous cellular gene to at least about 150%.

63. A method of modulating expression of an endogenous cellular gene in a cell, the method comprising the step of:

contacting a first target site in the endogenous cellular gene with a first zinc finger protein;

thereby modulating expression of the endogenous cellular gene.

64. The method of claim 63, wherein the step of contacting further comprises contacting a second target site in the endogenous cellular gene with a second zinc finger protein.

65. The method of claim 64, wherein the first and second target sites are adjacent.

66. The method of claim 65, wherein the first and second zinc finger proteins are covalently linked.

67. The method of claim 63, wherein the first zinc finger protein is a fusion protein comprising a regulatory domain.

68. The method of claim 67, wherein the first zinc finger protein is a fusion protein comprising at least two regulatory domains.

69. The method of claim 64, wherein the first and second zinc finger proteins are fusion proteins, each comprising a regulatory domain.

70. The method of claim 69, wherein the first and second zinc finger protein are fusion proteins, each comprising at least two regulatory domains.

71. The method of claim 63, wherein the cell is selected from the group consisting of animal cell, a plant cell, a bacterial cell, a protozoal cell, or a fungal cell.

promoter, and wherein the method further comprises the step of first administering the expression vector to the cell.

52. The method of claim 51, wherein the expression vector is a viral expression vector.

53. The method of claim 51, wherein the expression vector is a retroviral expression vector, an adenoviral vector, or an AAV expression vector.

54. The method of claim 51, wherein the zinc finger protein is encoded by a nucleic acid operably linked to an inducible promoter.

55. The method of claim 51, wherein the zinc finger protein is encoded by a nucleic acid operably linked to a weak promoter.

56. The method of claim 31, wherein the cell comprises less than about $1.5x10^6$ copies of the zinc finger protein.

57. The method of claim 31, wherein the target site is upstream of a transcription initiation site of the endogenous cellular gene.

58. The method of claim 31, wherein the target site is adjacent to a transcription initiation site of the endogenous cellular gene.

59. The method of claim 31, wherein the target site is adjacent to an RNA polymerase pause site downstream of a transcription initiation site of the endogenous cellular gene.

60. The method of claim 31, wherein the zinc finger protein comprises an SP-1 backbone.

61. The method of claim 60, wherein the zinc finger protein comprises a regulatory domain and is humanized.

62. A method of activating expression of an endogenous cellular gene, the method comprising the step of:

41. The method of claim 31, wherein the cell is selected from the group consisting of an animal cell, a plant cell, a bacterial cell, a protozoal cell, or a fungal cell.

42. The method of claim 41, wherein the cell is a mammalian cell.

43. The method of claim 42, wherein the cell is a human cell

44. The method of claim 31, wherein expression of the endogenous cellular gene is activated to at least about 200-500%.

45. The method of claim 31, wherein the endogenous cellular gene is a selected from the group consisting of FAD2-1, EPO, GM-CSF, GDNF, VEGF, and LDL-R.

46. The method of claim 31, wherein the endogenous cellular gene is VEGF.

47. The method of claim 31, wherein the activation of gene expression prevents repression of gene expression.

48. The method of claim 37 or 39, wherein the regulatory domain is selected from the group consisting of a transcriptional activator, or a histone acetyltransferase.

49. The method of claim 31, wherein the method further comprises the step of first administering to the cell a delivery vehicle comprising the zinc finger protein, wherein the delivery vehicle comprises a liposome or a membrane translocation polypeptide.

50. The method of claim 31, wherein the zinc finger protein is encoded by a zinc finger protein nucleic acid operably linked to a promoter, and wherein the method further comprises the step of first administering the nucleic acid to the cell in a lipid:nucleic acid complex or as naked nucleic acid.

51. The method of claim 31, wherein the zinc finger protein is encoded by an expression vector comprising a zinc finger protein nucleic acid operably linked to a

31. A method of activating expression of an endogenous cellular gene, the method comprising the step of:

contacting a first target site in the endogenous cellular gene with a first zinc finger protein, wherein the K_d of the zinc finger protein is less than about 25 nM;

thereby activating expression of the endogenous cellular gene to at least about 150%.

32. The method of claim 31, wherein the endogenous cellular gene is developmentally silent or inactive.

33. The method of claim 32, wherein the endogenous cellular gene is EPO, GATA, hemoglobin gamma, hemoglobin delta, an interleukin, GM-CSF, eutrophin, or MyoD.

34. The method of claim 31, wherein the step of contacting further comprises contacting a second target site in the endogenous cellular gene with a second zinc finger protein.

35. The method of claim 34, wherein the first and second target sites are adjacent.

36. The method of claim 35, wherein the first and second zinc finger proteins are covalently linked.

37. The method of claim 31, wherein the first zinc finger protein is a fusion protein comprising a regulatory domain.

38. The method of claim 37, wherein the first zinc finger protein is a fusion protein comprising at least two regulatory domains.

39. The method of claim 34, wherein the first and second zinc finger proteins are fusion proteins, each comprising a regulatory domain.

40. The method of claim 39, wherein the first and the second zinc finger protein are fusion proteins, each comprising at least two regulatory domains.

21. The method of claim 19, wherein the expression vector is a retroviral expression vector, an adenoviral expression vector, or an AAV expression vector.

22. The method of claim 19, wherein the zinc finger protein is encoded by a nucleic acid operably linked to an inducible promoter.

23. The method of claim 19, wherein the zinc finger protein is encoded by a nucleic acid operably linked to a weak promoter.

24. The method of claim 1, wherein the cell comprises less than about $1.5x10^6$ copies of the zinc finger protein.

25. The method of claim 1, wherein the target site is upstream of a transcription initiation site of the endogenous cellular gene.

26. The method of claim 1, wherein the target site is adjacent to a transcription initiation site of the endogenous cellular gene.

27. The method of claim 1, wherein the target site is adjacent to an RNA polymerase pause site downstream of a transcription initiation site of the endogenous cellular gene.

28. The method of claim 1, wherein the zinc finger protein comprises an SP-1 backbone.

29. The method of claim 28, wherein the zinc finger protein comprises a regulatory domain and is humanized.

30. A method of inhibiting expression of an endogenous cellular gene in a cell, the method comprising the step of:

contacting a target site in the endogenous cellular gene with a fusion zinc finger protein comprising six fingers and a regulatory domain, wherein the K_d of the zinc finger protein is less than about 25 nM;

thereby inhibiting expression of the endogenous cellular gene by at least about 20%.

12. The method of claim 1, wherein expression of the endogenous cellular gene is inhibited by at least about 75%-100%.

13. The method of claim 1, wherein the endogenous cellular gene is a selected from the group consisting of VEGF, ERα, IGF-I, c-myc, c-myb, ICAM, and Her2/Neu.

14. The method of claim 1, wherein the endogenous cellular gene is VEGF.

15. The method of claim 1, wherein the inhibition of gene expression prevents gene activation.

16. The method of claim 5 or 7, wherein the regulatory domain is selected from the group consisting of a transcriptional repressor, an endonuclease, a methyl transferase, and a histone deacetylase.

17. The method of claim 1, wherein the method further comprises the step of first administering to the cell a delivery vehicle comprising the zinc finger protein, wherein the delivery vehicle comprises a liposome or a membrane translocation polypeptide.

18. The method of claim 1, wherein the zinc finger protein is encoded by a zinc finger protein nucleic acid operably linked to a promoter, and wherein the method further comprises the step of first administering the nucleic acid to the cell in a lipid:nucleic acid complex or as naked nucleic acid.

19. The method of claim 1, wherein the zinc finger protein is encoded by an expression vector comprising a zinc finger protein nucleic acid operably linked to a promoter, and wherein the method further comprises the step of first administering the expression vector to the cell.

20. The method of claim 19, wherein the expression vector is a viral expression vector.

WHAT IS CLAIMED IS:

1. A method of inhibiting expression of an endogenous cellular gene in a cell, the method comprising the step of:

contacting a first target site in the endogenous cellular gene with a first zinc finger protein, wherein the K_d of the zinc finger protein is less than about 25 nM;

thereby inhibiting expression of the endogenous cellular gene by at least about 20%.

2. The method of claim 1, wherein the step of contacting further comprises contacting a second target site in the endogenous cellular gene with a second zinc finger protein.

3. The method of claim 2, wherein the first and second target sites are adjacent.

4. The method of claim 3, wherein the first and second zinc finger proteins are covalently linked.

5. The method of claim 1, wherein the first zinc finger protein is a fusion protein comprising a regulatory domain.

6. The method of claim 5, wherein the first zinc finger protein is a fusion protein comprising at least two regulatory domains.

7. The method of claim 2, wherein the first and second zinc finger proteins are fusion proteins, each comprising a regulatory domain.

8. The method of claim 7, wherein the first and second zinc finger protein are fusion proteins, each comprising at least two regulatory domains.

9. The method of claim 1, wherein the cell is selected from the group consisting of animal cell, a plant cell, a bacterial cell, a protozoal cell, or a fungal cell.

10. The method of claim 9, wherein the cell is a mammalian cell

11. The method of claim 10, wherein the cell is a human cell

transactivation domain, as described herein and in USSN 09/229,007, filed January 12, 1999. All plasmid DNAs were prepared using the Qiagen Midi preparation system.

$2x10^6$ Hep3B cells (a human hepatocellular carcinoma-derived cell line) or $5x10^6$ HEK293 cells (a human embryonic kidney epithelium-derived cell line) were seeded into 6-well plates one day before transfection. 500 ng of the effector plasmid (encoding the engineered ZFP) was transiently transfected into the cells using Lipofectamin (GIBCO-BRL). Mock transfection and transfection with an empty expression vector served as controls. One day later the growth medium was removed, and fresh DMEM was added. Culture supernatants were collected 24 hours later for determination of EPO protein expression levels using a commercially available ELISA kit (R&D Systems).

The results in Figure 11a show that transfection of a vector encoding the EPO2C ZFP transactivation protein significantly increased the level of EPO expression when compared to control vector (pcDNANVF) or mock transfected cells. This activation was observed in both Hep3B and HEK293 cells (Figure 11a).

Hep3B cells are known to be capable of expressing the EPO gene under certain conditions (Goldberg *et al., PNAS* 84:7972 (1987)). These cells are derived from hepatocytes, the fetal source of EPO. In Hep3B cells, hypoxia (or treatment with $CoCl_2$, which mimics hypoxia) activates expression of the EPO gene, which is otherwise silent (*see* Figure 11a). VEGF expression is also controlled by hypoxia in Hep3B cells.

Hypoxia or $CoCl_2$ treatment activates VEGF expression in HEK293 cells, showing that the hypoxic response functions normally in this cell line (*see* Figure 8). However, hypoxia or $CoCl_2$ treatment is not able to induce EPO expression in HEK293 cells (Figure 11a). This result occurs because HEK293 cells are derived from embryonic kidney epithelial cells, a tissue that does not act as a source of EPO. Despite inactivation of EPO in these cells, the chimeric EPO2C-VP16 ZFP is able to activate the EPO gene (Figure 11a).

The $CoCl_2$ effects and ZFP effects on EPO gene expression were shown to be effects on EPO transcription. Quantitative RT-PCR was performed using a method known as a TaqMan assay (PE Applied Biosystems). Data are shown in Figure 11b.

10, panel A; *see* Example VII for experimental details). The size of the protected probe was identical to the size of the probe generated from the control human RNA provided as a control for RNA integrity. (Figure 10, panel A).

In a separate experiment, the level of VEGF specific mRNA was also quantitated in cells that had been transfected with a VEGF-KRAB effector plasmid (Figure 10, panel B; *see* Example VI for experimental details). The details of the transfection are described in Example VI. A dramatic decrease in the level of VEGF mRNA was observed when cells were transfected with the VEGF3a/1-KRAB effector plasmid. No significant decrease in VEGF mRNA was observed when cells were transfected with NKF-control or a non-VEGF specific ZFP (CCR5-5-KRAB and CCR5-3-KRAB, which recognize different CCR5 target sites).

This experiment demonstrates that the increase in VEGF protein observed upon transfection with the VEGF1-VP16 chimeric transcription factor is mediated by an increase in the level of VEGF mRNA. Similarly, the decrease in VEGF protein observed upon transfection with the VEGF3a/1-KRAB chimeric transcription factor is mediated by a decrease in the level of VEGF mRNA.

Example IX: Activation of EPO, a developmentally silent gene

EPO is a 30.4 kD glycoprotein hormone and plays a key role in the control of red blood cell production. The EPO gene is normally only expressed in fetal liver and the adult kidney in response to hypoxia. It acts through interacting with the EPO receptor on the erythroid progenitor cells in the bone marrow to stimulate their proliferation and differentiation into mature erythrocytes (Jelkmann, *Physiological Reviews* 72:449(1992); Semenza, *Hematology/Oncology Clinics of North America* 8:863(1994); Ratcliffe *et al.*, *J. Exp. Bio.*(1991)). Recombinant human EPO gene products have been successfully used to treat anemia and chronic renal failure (Egrie, *Pharacotherapy* 10:3S(1990)).

The EPO2C ZFP was designed to recognize a 9-bp DNA-binding site located 853-bp upstream of the EPO transcription initiation site. Methods for design and construction of EPO2C are described herein and in USSN 09/229,007, filed January 12, 1999. The EPO2C binding site sequence is GCGGTGGCT.

Eukaryotic expression vectors were constructed by fusing the sequence encoding EPO2C ZFP to the SV40 nuclear localization signal (NLS) and the HSV VP16

DNA-Lipofectamine mixture was removed, and fresh culture medium containing 10% fetal bovine serum was layered on the cells. One day later, fresh media was added and the supernatant was collected 24 hours later for determination of VEGF levels using a commercially available ELISA kit (R and D Systems).

For the three-fingered VEGF1-specific ZFP (VEGF1-VP16), a 7-10 fold increase in VEGF expression was observed when compared to control plasmid (NVF-control) and mock transfected cells (Figure 9). Similar results have been obtained in 5 independent experiments. It is important to note that the level of VEGF secretion in VEGF1-VP16 transfected cells was equivalent or greater than the level in cells that have been treated with DFX (Figure 9). Introduction of VEGF3a/1-VP16 stimulated a more modest induction of VEGF. This result is consistent with the finding in Example VI, in which expression of the 18-bp binding protein without a functional domain prevented activation to a certain degree. This result suggested that the tight binding of this protein to the start site of transcription interferes with activation.

These data indicate that a designed ZFP is capable of locating and binding to its target site on the chromosome, presenting a transcriptional activation domain, and dramatically enhancing the expression level of that gene. In particular, the results indicate that ZFPs with a K_d of less than about 25 nM (e.g., VEGF1 has an average apparent K_d of about 10 nM) provide dramatic increases in expression.

Example VIII: RNase protection assay

To further substantiate the results in Examples VI and VII, a ribonuclease protection assay (RPA) was performed to correlate the increased level of VEGF protein with an increase in VEGF mRNA levels (Example VII) , and to correlate the decreased level of VEGF protein with a decrease in VEGF mRNA levels (Example VI).

RNA was isolated from the transfected cells using an RNA isolation kit (Pharmingen). Radiolabeled multi template probes, which included a VEGF specific probe, were prepared by *in vitro* transcription and hybridized overnight at 56^0C to 5 μg of each of the RNAs from the experimental and control transfected cells. The hybridization mixture was treated with RNase and the protected probes were purified and subjected to 5% denaturing polyacrylamide gel electrophoresis and the radioactivity was evaluated by autoradiography. 293 cells transfected with the VEGF1-VP16 had a 2-4 fold increase in the level of VEGF mRNA when compared to cells transfected with NVF-control (Figure

expression was observed with VEGF3a/1-NF, which expresses the 18-bp binding protein without a functional domain. This result suggests that binding to the start site of transcription, even without a repression domain, interferes with transcription. Even when fused to the KRAB domain, the VEGF3a ZFP is unable to affect expression levels (plasmid VEGF3a-KRAB). However, VEGF1 fused to KRAB (VEGF1-KRAB) results in a dramatic decrease in expression. VEGF3a/1 fused to KRAB (VEGF3a/1-KRAB) prevents expression of VEGF altogether.

These data indicate that a designed ZFP is capable of locating and binding to its target site on the chromosome and preventing expression of an endogenous cellular target gene. In particular, the results indicate that ZFPs with a K_d of less than about 25 nM (e.g., VEGF1 has an average apparent K_d of about 10 nM) provide dramatic decreases in expression. In addition, the data demonstrate that the KRAB functional domain enhances gene silencing. Because in this experiment the introduction of the repressor occurs before the inducer of VEGF is added (DFX), the data demonstrate the ability of a designed repressor to prevent activation of an already quiescent gene. In addition, these results demonstrate that a six-finger engineered ZFP (VEGF3a/1) with nanomolar affinity for its target is able to inhibit the hypoxic response of the VEGF gene when it binds a target that overlaps the transcriptional start site.

Example VII: Activation of endogenous VEGF gene in human cells

This Example demonstrates that a designed ZFP can activate the expression of a gene that is in its natural context and chromatin structure. Specifically, effector plasmids expressing VEGF ZFPs fused to the VP16 activation domain were introduced into cells and were shown to up-regulate the VEGF gene.

Eucaryotic expression vectors were constructed that fuse the VEGF3a/1 and the VEGF1 ZFPs to the SV40 NLS and VP16, as described in Example III. Transfections were done using Lipofectamine, a commercially available liposome preparation from GIBCO-BRL. All plasmid DNAs were prepared using the Qiagen Midi DNA purification system. 10 μg of the effector plasmid (containing the engineered ZFP) was mixed with 100 μg of Lipofectamine (50 μl) in a total volume of 1600μl of Opti-MEM. A pCMVβ-gal plasmid (Promega) was also included in the DNA mixture as an internal control for transfection efficiency. Following a 30 minute incubation, 6.4 ml of DMEM was added and the mixture was layered on 3 x10^6 293 cells. After five hours, the

efficiency. Following a 30 minute incubation, 6.4 ml of DMEM was added and the mixture was layered on 3 x10^6 293 cells. After five hours, the DNA-Lipofectamine mixture was removed, and fresh culture medium containing 10% fetal bovine serum was layered on the cells.

Eighteen hours post transfection, the 293 cells were induced by treatment with 100 μM DFX (desferrioxamine), resulting in a rapid and lasting transcriptional activation of the VEGF gene and also in a gradual increase in VEGF mRNA stability (Ikeda *et al., J. Biol. Chem.* 270:19761-19766 (1995)). Under routine culture conditions, 293 cells secrete a low level of VEGF in the culture media. The cells were allowed to incubate an additional 24 hours before the supernatants were collected for determination of VEGF levels by an ELISA assay.

In parallel experiments that demonstrated a similar level of repression, cell viability was monitored using the Promega Celltiter 96® Aqueous One Solution cell proliferation assay (Promega). After Dfx treatment for 18 hours, 500 μL of the original 2 ml of media was removed and analyzed for VEGF expression, as described above. To evaluation cell viability, 300 μL of Promega Celltiter 96® Aqueous One Solution Reagent was added to the remaining 1.5 ml. The cells were then incubated at 37°C for approximately 2 hours. 100 μL from each well was transferred to a 96-well plate and read on an ELISA plate reader at OD 490 nm. There was no significant reduction in viability of cells expressing the VEGF3a/1-KRAB construct relative to those transfected with empty vector controls, indicating that the VEGF repression observed was not due to generalized cell death.

A 40-50-fold decrease in VEGF expression was noted in the DFX treated cells transfected with VEGF3a/1-KRAB, an expression vector encoding the 18 bp binding VEGF high affinity ZFP. A two-fold decrease in expression was observed when cells were transfected with VEGF1-KRAB, an expression vector encoding the 9 bp binding VEGF high affinity ZFP. No significant decrease in VEGF expression was observed in cells that were transfected with a non-VEGF ZFP (CCR5-KRAB) or NKF-control (Figure 8). Similar results have been obtained in three independent transfection experiments.

In a separate experiment, the following results were obtained (data not shown). VEGF1-NF, which expresses the 9-bp-binding VEGF1 ZFP without a functional domain, showed no effect on VEGF gene expression. A significant reduction in VEGF

activation relative to the empty pcDNA vector control. VEGF3a/1 (the 18-bp-binding ZFP) expression plasmid activates luciferase expression very substantially, showing about a 14-fold increase relative to pcDNA. These experiments clearly demonstrate that a designed ZFP, when fused to the VP16 activation domain, is capable of functioning in a cell to activate transcription of a gene when its target site is present. Furthermore, these results clearly demonstrate that an 18-bp binding protein, VEGF3a/1, is a much better activator in this assay than a 9-bp binding VEGF1 protein. This could be a result of the improved affinity or decreased off-rate of the VEGF3a/1 protein.

A fourth VEGF reporter plasmid was constructed by cloning the KpnI/NcoI fragment of pVFR2-4x into pGL3-Promoter to create plasmid pVFR4-4x. Activation was observed in co-transfections using this reporter in combination with effector plasmids expressing the VEGF1-VP16 and VEGF3a/1-VP16 fusions (data not shown). This indicates that these artificial trans-activators are functional when bound either upstream or downstream of the start of transcription.

These co-transfection data demonstrate that ZFPs can be used to regulate expression of reporter genes. Such experiments serve as a useful tool for identifying ZFPs for further use as modulators of expression of endogenous cellular genes. As is shown below, modulation results can vary between co-transfection experiments and endogenous gene experiments, while using the same ZFP construct.

Example VI: Repression of an endogenous VEGF gene in human cells

This Example demonstrates that a designed ZFP can repress expression of an endogenous cellular gene that is in its natural context and chromatin structure. Specifically, effector plasmids expressing VEGF ZFPs fused to the KRAB repression domain were introduced into cells and were shown to down-regulate the VEGF gene.

Eucaryotic expression vectors were constructed that fuse the VEGF3a/1 and the VEGF1 ZFPs to the SV40 NLS and KRAB, as described above in Example III. Transfections were done using Lipofectamine, a commercially available liposome preparation from GIBCO-BRL. All plasmid DNAs were prepared using Qiagen Midi DNA purification system. 10 μg of the effector plasmid was mixed with 100 μg of Lipofectamine (50 μl) in a total volume of 1600 μl of Opti-MEM. A pCMVβ-gal plasmid (Promega) was also included in the DNA mixture as an internal control for transfection

above, at the MluI/BglII sites of plasmid pGL3-Promoter (Promega). This vector has been deleted for the SV40 enhancer sequence and therefore has a lower basal level of firefly luciferase expression. pVFR3-4x was constructed by swapping the KpnI/NcoI fragment of pVFR1-4x into the KpnI/NcoI sites of pGL3-Promoter.

The effector plasmid construction is described above. The VEGF1-VP16, VEGF3a-VP16, and VEGF3a/1-VP16 expression vectors were designed to produce a fusion of the SV40 nuclear localization sequence, the VEGF ZFP, the VP16 trans-activation domain, and a FLAG epitope tag all under the control of the CMV promoter. The empty pcDNA3 expression vector was used as a control.

All vectors were prepared using Qiagen DNA purification kits. Figure 7 shows a typical set of transfections using 293 (human embryonic kidney) cells. Approximately 40,000 cells were seeded into each well of a 24-well plate and allowed to grow overnight in D-MEM medium containing 10% fetal bovine serum at 37°C with 5% CO_2. Cells were washed with serum-free D-MEM and overlayed with 200 µl of the same. Plasmids were introduced using a calcium phosphate transfection kit (Gibco-BRL) according to the manufacturer's instructions. Cells in each well were transfected with 1.5 µg of reporter plasmid, 1.5 µg of effector plasmid, and 0.5 µg of an actin/β-gal plasmid. Plasmids were combined with 15 µl of $CaCl_2$ and brought to 100 µl with dH_2O. 100 µl of HEPES solution was added dropwise while vortexing. The mix was incubated for 30 min at room temperature. The 200 µl of calcium phosphate-treated DNA was then added to the medium in each well. Transfections were done in triplicate. After 5 hours, the medium was removed and 1 ml of medium containing 10% serum was added. Cells were harvested 40-48 hours after transfection. Luciferase assays were done using the Dual-Light™ system (Tropix). The third plasmid transfected, actin/β-gal, carries the β-galactosidase gene under the control of the actin promoter and was co-transfected as a standard for transfection efficiency. The β-galactosidase assays were also done according to the manufacturer's protocol (Tropix). The data shown in Figure 7 are the average of triplicate assays normalized against the β-galactosidase activity.

For the control reporter plasmid, pGL3-Promoter (pGL3-P), the presence or absence of the ZFP-VP16 expression plasmid does not significantly influence the luciferase expression level. For pVFR3-4x, the reporter containing four copies of the VEGF target site, presence of VEGF1 (the 9-bp-binding ZFP) shows a very slight

assays were done using the Dual Luciferase™ System (Promega). The third plasmid transfected, pRL-SV40, carries the *Renilla* luciferase gene and was co-transfected as a standard for transfection efficiency. The data shown in Figure 6 are the averages of triplicate assays normalized against the *Renilla* activity.

For the control reporter plasmid pGL3-Control (pGL3-C), the presence or absence of the ZFP-KRAB expression plasmid does not influence the luciferase expression level. However, for pVFR1-4x, the reporter containing four copies of the VEGF target site, presence of the VEGF1 (9-bp-binding ZFP) or VEGF3a/1 (18-bp-binding ZFP) expression plasmid reduces luciferase expression by a factor of 2-3 relative to the empty pcDNA vector control. The VEGF3a (9-bp-binding ZFP) expression plasmid appears to exhibit little or no effect. These experiments clearly demonstrate that a designed ZFP is capable of functioning in a cell to repress transcription of a gene when its target site is present. Furthermore, it appears that a certain level of affinity is required for function; i.e., VEGF1 and VEGF3a/1, with K_ds of 10 nM or less, are functional, whereas VEGF3a, with a K_d of 200 nM, is not.

A second reporter plasmid, pVFR2-4x, was constructed by removing the four copies of the VEGF target sites using HindIII and inserted them into the HindIII site of pGL3-Control (in the forward orientation). This places the target sites between the start site of transcription for the SV40 promoter and the translational start codon of the luciferase gene. In similar co-transfection experiments to those described, approximately 3-4 fold repression of the luciferase signal was observed with the VEGF1-KRAB or VEGF3a/1-KRAB repressors relative to the pcDNA controls (data not shown). This indicates that the repressors are active when bound either upstream or downstream of the start of transcription.

Example V: Activation of VEGF reporters in co-transfection experiments

This Example demonstrates the use of transient co-transfection studies to measure the activity of the ZFP transcriptional activators in cells. The experimental setup is similar to that of Example IV except that a different transfection method, a different cell line, and a different set of reporter and effector plasmids was used.

For activation experiments, a reporter was constructed labeled pVFR3-4x. This reporter contains the four copies of the VEGF targets, with the sequence shown

The reporter plasmid system was based on the pGL3 firefly luciferase vectors (Promega). Four copies of the VEGF target sites were inserted upstream of the SV40 promoter, which is driving the firefly luciferase gene, in the plasmid pGL3-Control to create pVFR1-4x. This plasmid contains the SV40 enhancer and expresses firefly luciferase to high levels in many cell types. Insertions were made by ligating together tandem copies of the two complementary 42-bp oligonucleotides, JVF9 and JVF10, described in Example II. Adaptor sequences were ligated on, and the assembly was inserted into the MluI/BglII sites of pGL3-Control. This resulted in the insertion of the following sequence between those sites:

ACGCGTaagcttGCTAGCGAGCGGGGAGGATCGCGGAGGCTTGGGGCAGCCGGG
TAGAGCGAGCGGGGAGGATCGCGGAGGCTTGGGGCAGCCGGGTAGAGCGAG
CGGGGAGGATCGCGGAGGCTTGGGGCAGCCGGGTAGAGCGAGCGGGGAGGA
TCGCGGAGGCTTGGGGCAGCCGGGTAGAGCGCTCAGaagcttAGATCT.

The first six and last six nucleotides shown are the MluI and BglII sites; the lowercase letters indicate HindIII sites. The binding sites for VEGF1 and VEGF3a are underlined.

The effector plasmid construction is described above. The VEGF1-KRAB, VEGF3a-KRAB, and VEGF3a/1-KRAB expression vectors were designed to produce a fusion of the SV40 nuclear localization sequence, the VEGF ZFP, the KRAB repression domain, and a FLAG epitope marker all under the control of the CMV promoter. The empty pcDNA3.1 expression vector was used as a control (pcDNA).

All vectors were prepared using Qiagen DNA purification kits. Figure 6 shows a typical set of transfections using COS-1 (African green monkey kidney) cells. Approximately 40,000 cells were seeded into each well of a 24-well plate and allowed to grow overnight in Dulbecco's Modified Eagle Medium (D-MEM) medium containing 10% fetal bovine serum at 37°C with 5% CO_2. Cells were washed with PBS and overlayed with 200 µl of serum-free D-MEM. Plasmids were introduced using lipofectamine (Gibco-BRL). Each well was transfected with about 0.3 µg of effector plasmid, 0.3 µg of reporter plasmid, and 0.01 µg of plasmid pRL-SV40 (Promega) that had been complexed with 6 µl of lipofectamine and 25 µl of D-MEM for 30 min at 37°C. Transfections were done in triplicate. After 3 hrs, 1 ml of medium containing 10% serum was added to each well. Cells were harvested 40-48 hours after transfection. Luciferase

ATGGGGTACCCGGGGATGGATCCGGCAGCGACTACAAGGACGACGATGACA
AGTAAGCTTCTCGAG

into the EcoRI-XhoI sites of pcDNA-NKF, thereby replacing the NLS-KRAB-FLAG sequences with NLS-FLAG only.

VEGF1-NF and VEGF3a/1-NF were constructed by inserting a KpnI-BamHI cassette containing the ZFP sequences into NF-control digested with KpnI and BamHI. CCR5-KRAB was constructed in the same way as the VEGF KRAB vectors, except that the ZFP sequences were designed to be specific for a DNA target site that is unrelated to the VEGF targets.

Finally, control versions of both the KRAB and VP16 expression plasmids were constructed. Plasmid NKF-control was designed to express NLS-KRAB-FLAG without zinc finger protein sequences; plasmid NVF-control was designed to express NLS-VP16-FLAG without ZFP sequences. These plasmids were made by digesting pcDNA-NKF and -NVF, respectively, with BamHI, filling in the ends with Klenow, and religating in order to place the downstream domains into the proper reading frame. These plasmids serve as rigorous controls for cell culture studies.

Mammalian cell expression and nuclear localization of the VEGF engineered ZFPs was demonstrated through immunofluorescence studies. 293 (human embryonic kidney) cells were transfected with the expression plasmid encoding the NLS-VEGF1-KRAB-FLAG chimera. Lipofectamine was used as described below. After 24-48 hours, cells were fixed and exposed to a primary antibody against the FLAG epitope. A secondary antibody labeled with Texas Red was applied, and the cells were counter stained with DAPI. Texas Red staining was observed to consistently co-localize with the DAPI staining, indicating that the ZFP being expressed from this plasmid was nuclear localized.

Example IV: Repression of VEGF reporters in co-transfection experiments

This Example demonstrates the use of transient co-transfection studies to measure the activity of the ZFP repressor proteins in cells. Such experiments involve co-transfection of ZFP-KRAB expression ("effector") plasmids with reporter plasmids carrying the VEGF target sites. Efficacy is assessed by the repression of reporter gene expression in the presence of the effector plasmid relative to empty vector controls.

Thiesen, 1990, *supra*), the FLAG epitope (from Kodak/IBI catalog), and a HindIII site, altogether comprising the sequence

GGTACCCGGGGATCCCGGACACTGGTGACCTTCAAGGATGTATTTGTGGACTT
CACCAGGGAGGAGTGGAAGCTGCTGGACACTGCTCAGCAGATCGTGTACAGA
AATGTGATGCTGGAGAACTATAAGAACCTGGTTTCCTTGGGCAGCGACTACA
AGGACGACGATGACAAGTAAGCTTCTCGAG

where the KpnI, BamHI and XhoI sites are underlined.

The VEGF3a/1-KRAB effector plasmid was generated by inserting a KpnI-BamHI cassette containing the ZFP sequences into pcDNA-NKF digested with KpnI and BamHI. The VEGF1-KRAB and VEGF3a-KRAB effector plasmids were constructed in a similar way except that the ZFP sequences were first cloned into the NLS-KRAB-FLAG sequences in the context of plasmid pLitmus 28 (New England Biolabs) and subsequently moved to the BamHI-XhoI sites of pcDNA3.1(+) as a BglII-XhoI cassette, where the BglII site was placed immediately upstream of the EcoRI site (see Example IV for expression of these vectors).

The effector plasmids used in Example V were constructed as follows. Plasmid pcDNA-NVF was constructed by PCR amplifying the VP16 transactivation domain, as described above, and inserting the product into the BamHI/HindIII sites of pcDNA-NKF, replacing the KRAB sequence. The sequence of the inserted fragment, from BamHI to HindIII, was:

GGATCCGCCCCCCCGACCGATGTCAGCCTGGGGGACGAGCTCCACTTAGACG
GCGAGGACGTGGCGATGGCGCATGCCGACGCGCTAGACGATTTCGATCTGGA
CATGTTGGGGGACGGGGATTCCCCGGGGCCGGGATTTACCCCCCACGACTCC
GCCCCCTACGGCGCTCTGGATATGGCCGACTTCGAGTTTGAGCAGATGTTTAC
CGATGCCCTTGGAATTGACGAGTACGGTGGGGGCAGCGACTACAAGGACGAC
GATGACAAGTAAGCTT.

VEGF1-VP16 and VEGF3a/1-VP16 vectors were constructed by inserting a KpnI-BamHI cassette containing the ZFP sequences into pcDNA-NVF digested with KpnI and BamHI.

The effector plasmids used in Example VI were constructed as follows. Plasmid NF-control was generated by inserting the sequence

GAATTCGCTAGCGCCACCATGGCCCCCAAGAAGAAGAGGAAGGTGGGAATCC

CGCGGATCCGCCCCCCCGACCGATG, and

(2) JVF25

CCGCAAGCTTACTTGTCATCGTCGTCCTTGTAGTCGCTGCCCCCACCGTACTC GTCAATTCC.

The downstream primer, JVF25, was designed to include a downstream FLAG epitope-encoding sequence.

Three expression vectors were constructed for these studies. The general design is summarized in Figure 5. The vectors are derived from pcDNA3.1(+) (Invitrogen), and place the ZFP constructs under the control of the cytomegalovirus (CMV) promoter. The vector carries ampicillin and neomycin markers for selection in bacteria and mammalian cell culture, respectively. A Kozak sequence for proper translation initiation (Kozak, *J. Biol. Chem.* 266:19867-19870 (1991)) was incorporated. To achieve nuclear localization of the products, the nuclear localization sequence (NLS) from the SV40 large T antigen (Pro-Lys-Lys-Lys-Arg-Lys-Val) (Kalderon *et al., Cell* 39:499-509 (1984)) was added. The insertion site for the ZFP-encoding sequence is followed by the functional domain sequence. The three versions of this vector differ in the functional domain; "pcDNA-NKF" carries the KRAB repression domain sequence, "pcDNA-NVF" carries the VP16 activation domain, and "NF-control" carries no functional domain. Following the functional domain is the FLAG epitope sequence (Kodak) to allow specific detection of the ZFPs.

The vectors were constructed as follows. Plasmid pcDNA-ΔHB was constructed by digesting plasmid pcDNA3.1(+) (Invitrogen) with HindIII and BamHI, filling in the sticky ends with Klenow, and religating. This eliminated the HindIII, KpnI, and BamHI sites in the polylinker. The vector pcDNA3.1(+) is described in the Invitrogen catalog. Plasmid pcDNA-NKF was generated by inserting a fragment into the EcoRI/XhoI sites of pcDNA-ΔHB that contained the following: 1) a segment from EcoRI to KpnI containing the Kozak sequence including the initiation codon and the SV40 NLS sequence, altogether comprising the DNA sequence

GAATTCGCTAGCGCCACCATGGCCCCCAAGAAGAAGAGGAAGGTGGGAATCC ATGGGGTAC,

where the EcoRI and KpnI sites are underlined; and 2) a segment from KpnI to XhoI containing a BamHI site, the KRAB-A box from KOX1 (amino acid coordinates 11-53 in

occupancy of the target site is much higher for the 18-bp binding protein than for the 9-bp binding protein.

Example III: Fusing designed ZFP sequences to functional domains in mammalian expression vectors

This Example describes the development of expression vectors for producing ZFPs within mammalian cells, translocating them to the nucleus, and providing functional domains that are localized to the target DNA sequence by the ZFP. The functional domains employed are the Kruppel-Associated Box (KRAB) repression domain and the Herpes Simplex Virus (HSV-1) VP16 activation domain.

Certain DNA-binding proteins contain separable domains that function as transcriptional repressors. Approximately 20% of ZFPs contain a non-DNA-binding domain of about 90 amino acids that functions as a transcriptional repressor (Thiesen, *The New Biologist* 2:363-374 (1990); Margolin *et al.*, *PNAS* 91:4509-4513 (1994); Pengue *et al.*, (1994), *supra*; Witzgall *et al.*, (1994), *supra*). This domain, termed the KRAB domain, is modular and can be joined to other DNA-binding proteins to block expression of genes containing the target DNA sequence (Margolin *et al.*, (1994); Pengue *et al.*, (1994); Witzgall *et al.*, (1994), *supra*). The KRAB domain has no effect by itself; it needs to be tethered to a DNA sequence via a DNA-binding protein to function as a repressor. The KRAB domain has been shown to block transcription initiation and can function at a distance of up to at least 3 kb from the transcription start site. The KRAB domain from the human KOX-1 protein (Thiesen, *The New Biologist* 2:363-37 (1990)) was used for the studies described here. This 64 amino acid domain can be fused to ZFPs and has been shown to confer repression in cell culture (Liu *et al.*, *supra*).

The VP16 protein of HSV-1 has been studied extensively, and it has been shown that the C-terminal 78 amino acids can act as a trans-activation domain when fused to a DNA-binding domain (Hagmann *et al.*, *J. Virology* 71:5952-5962 (1997)). VP16 has also been shown to function at a distance and in an orientation-independent manner. For these studies, amino acids 413 to 490 in the VP16 protein sequence were used. DNA encoding this domain was PCR amplified from plasmid pMSVP16ΔC+119 using primers with the following sequences:

(1) JVF24

CHIQGCGKVYGTTSNLRRHLRWHTGERPFMCTWSYCGKRFTRSSNLQRHKRTH TGEKKFACPECPKRFMRSDHLSRHIKTHQNKKGGS

The 18-bp binding protein VEGF3a/1 was expressed in *E. coli* as an MBP fusion, purified by affinity chromatography, and tested in EMSA experiments as described in Example I. The target oligonucleotides were prepared as described and comprised the following complementary sequences:

(1) JVF9

AGCGAGCGGGGAGGATCGCGGAGGCTTGGGGCAGCCGGGTAG, and

(2) JVF10

CGCTCTACCCGGCTGCCCCAAGCCTCCGCGATCCTCCCCGCT.

For the EMSA studies, 20 µl binding reactions contained 10 fmole (0.5 nM) 5'-^{32}P-labeled double-stranded target DNA, 35 mM Tris HCl (pH 7.8), 100 mM KCl, 1 mM $MgCl_2$, 5 mM dithiothreitol, 10% glycerol, 20 µg/ml poly dI-dC, 200 µg/ml bovine serum albumin, and 25 µM $ZnCl_2$. Protein was added as one fifth volume from a 3-fold dilution series. Binding was allowed to proceed for 60 min at either room temperature or 37°C. Polyacrylamide gel electrophoresis was carried out at room temperature or 37°C using precast 10% or 10-20% Tris-HCl gels (BioRad) and standard Tris-Glycine running buffer. The room temperature assays yielded an apparent K_d for this VEGF3a/1 protein of approximately 1.5 nM. Thus, the 18-bp binding ZFP bound with high affinity to its target site. In a parallel experiment, VEGF1 protein was tested against its target using the oligonucleotides described in Example I, yielding an apparent K_d of approximately 2.5 nM. When binding and electrophoresis were performed at 37°C, the apparent K_d of VEGF3a/1 was approximately 9 nM when tested against the 18-bp target, compared to a K_d of 40 nM for VEGF1 tested against its target. This indicates that the difference in binding affinities is accentuated at the higher temperature.

The apparent K_d is a useful measure of the affinity of a protein for its DNA target. However, for a DNA binding site either *in vitro* or *in vivo*, its occupancy is determined to a large extent by the off-rate of the DNA-binding protein. This parameter can be measured by competition experiments as shown in Figure 4. The conditions for EMSA were as described above; binding and electrophoresis were performed at 37°C. These data indicate that the half-life of the protein-DNA complex is more than ten times longer for VEGF3a/1 than for VEGF1. Thus, under thes *in vitro* conditions, the

at the ends (restriction sites underlined). VEGF1 was PCR amplified using the primers SPEamp13 (5'- GGAGCCAAGGCTGTGGTAAAGTTTACGG) and SPEamp11 (5'- GGAGAAGCTTGGATCCTCATTATCCC) to generate StyI and HindIII restriction sites at the ends (restriction sites underlined). Using synthetic oligonucleotides, the following sequence was ligated between the XbaI and StyI sites, where XbaI and StyI are underlined: TCT AGA CAC ATC AAA ACC CAC CAG AAC AAG AAA GAC GGC GGT GGC AGC GGC AAA AAG AAA CAG CAC ATA TGT CAC ATC CAA GG. This introduced the linker sequence DGGGS between the two SP-1 domains. The ligation product was reamplified with primers SPE7 and SPEamp11 and cloned into pUC19 using the EcoRI and HindIII sites. The linked ZFP sequences were then amplified with primers

(1) GB19

GCCATGCCGGTACCCATACCTGGCAAGAAGAAGCAGCAC)

(2) GB10

CAGATCGGATCCACCCTTCTTATTCTGGTGGGT to introduce KpnI and BamHI sites for cloning into the modified pMAL-c2 expression vector as described above.

The nucleotide sequence of the designed, 6-finger ZFP VEGF3a/1 from KpnI to BamHI is:

GGTACCCATACCTGGCAAGAAGAAGCAGCACATCTGCCACATCCAGGGCTGT
GGTAAAGTTTACGGCCAGTCCTCCGACCTGCAGCGTCACCTGCGCTGGCACAC
CGGCGAGAGGCCTTTCATGTGTACCTGGTCCTACTGTGGTAAACGCTTCACAC
GTTCGTCAAACCTACAGAGGCACAAGCGTACACACACAGGTGAGAAGAAATT
TGCTTGCCCGGAGTGTCCGAAGCGCTTCATGCGAAGTGACGAGCTGTCTAGA
CACATCAAAACCCACCAGAACAAGAAAGACGGCGGTGGCAGCGGCAAAAAG
AAACAGCACATATGTCACATCCAAGGCTGTGGTAAAGTTTACGGCACAACCT
CAAATCTGCGTCGTCACCTGCGCTGGCACACCGGCGAGAGGCCTTTCATGTGT
ACCTGGTCCTACTGTGGTAAACGCTTCACCCGTTCGTCAAACCTGCAGCGTCA
CAAGCGTACCCACACCGGTGAGAAGAAATTTGCTTGCCCGGAGTGTCCGAAG
CGCTTCATGCGTAGTGACCACCTGTCCCGTCACATCAAGACCCACCAGAATAA
GAAGGGTGGATCC

The VEGF3a/1 amino acid translation (using single letter code) is:

VPIPGKKKQHICHIQGCGKVYGQSSDLQRHLRWHTGERPFMCTWSYCGKRFTRS
SNLQRHKRTHTGEKKFACPECPKRFMRSDELSRHIKTHQNKKDGGGSGKKKQHI

the MBP-fused and unfused versions of the proteins bound with similar affinities. K_ds were also determined under these conditions for MBP fusions of the wild-type Zif268 and SP-1 ZFPs, which yielded K_ds of 60 and 65 nM, respectively. These results are similar to binding constants reported in the literature for Zif268 of approximately 2-30 nM (*see, e.g.*, Jamieson *et al., Biochemistry* 33:5689-5695 (1994)). The K_ds for the synthetic VEGF ZFPs therefore compare very favorably with those determined for these naturally-occurring DNA-binding proteins.

In summary, this Example demonstrates the generation of two novel DNA-binding proteins directed to specific targets near the transcriptional start of the VEGF gene. These proteins bind with affinities similar to those of naturally-occurring transcription factors binding to their targets.

Example II: Linking ZFPs to bind an 18-bp target in the human VEGF gene

An important consideration in ZFP design is DNA target length. For random DNA, a sequence of n nucleotides would be expected to occur once every 0.5 x 4^n base-pairs. Thus, DNA-binding domains designed to recognize only 9 bp of DNA would find sites every 130,000 bp and could therefore bind to multiple locations in a complex genome (on the order of 20,000 sites in the human genome). 9-bp putative repressor-binding sequences have been chosen for VEGF in the 5' UTR where they might directly interfere with transcription. However, in case zinc finger domains that recognize 9-bp sites lack the necessary affinity or specificity when expressed inside cells, a larger domain was constructed to recognize 18 base-pairs by joining separate three-finger domains with a linker sequence to form a six-finger protein. This should ensure that the repressor specifically targets the appropriate sequence, particularly under conditions where only small amounts of the repressor are being produced. The 9-bp target sites in VEGF were chosen to be adjacent to one another so that the zinc fingers could be linked to recognize an 18-bp sequence. The linker DGGGS was chosen because it permits binding of ZFPs to two 9-bp sites that are separated by a one nucleotide gap, as is the case for the VEGF1 and VEGF3a sites (*see also* Liu *et al., PNAS* 5525-5530 (1997)).

The 6-finger VEGF3a/1 protein encoding sequence was generated as follows. VEGF3a was PCR amplified using the primers SPE7 (5'-GAGCAGAATTCGGCAAGAAGAAGCAGCAC) and SPEamp12 (5'-GTGGTCTAGACAGCTCGTCACTTCGC) to generate EcoRI and XbaI restriction sites

VEGF site 3, top: 5'-CATGCATATCGCGGAGGCTTGGCATCGAT

VEGF site 3, bottom: 5'-ATCGATGCCAAGCCTCCGCGATATGCATG

The VEGF DNA target sites are underlined. The 3 bp on either side of the 9 bp binding site was also derived from the actual VEGF DNA sequence. The top strand of each target site was labeled with polynucleotide kinase and γ-^{32}P dATP. Top and bottom strands were annealed in a reaction containing each oligonucleotide at 0.5 μM, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 50 mM NaCl. The mix was heated to 95°C for 5 min. and slow cooled to 30°C over 60 min. Duplex formation was confirmed by polyacrylamide gel electrophoresis. Free label and ssDNA remaining in the target preparations did not appear to interfere with the binding reactions.

Binding of the ZFPs to target oligonucleotides was performed by titrating protein against a fixed amount of duplex substrate. Twenty microliter binding reactions contained 10 fmole (0.5 nM) 5'-^{32}P-labeled double-stranded target DNA, 35 mM Tris HCl (pH 7.8), 100 mM KCl, 1 mM $MgCl_2$, 1 mM dithiothreitol, 10% glycerol, 20 μg/ml poly dI-dC (optionally), 200 μg/ml bovine serum albumin, and 25 μM $ZnCl_2$. Protein was added as one fifth volume from a dilution series made in 200 mM NaCl, 20 mM Tris (pH 7.5), 1 mM DTT. Binding was allowed to proceed for 30 min. at room temperature. Polyacrylamide gel electrophoresis was carried out at 4°C using precast 10% or 10-20% Tris-HCl gels (BioRad) and standard Tris-Glycine running buffer containing 0.1 mM $ZnCl_2$.

The results of a typical EMSA using an MBP fused ZFP are shown in Figure 3. In this case, a 3-fold dilution series of the MBP-VEGF1 protein was used. The shifted product was quantitated on a phosphorimager (Molecular Dynamics) and the relative signal (percent of plateau value) vs. the $\log_{10}$ of nM protein concentration was plotted. An apparent K_d was found by determining the protein concentration that gave half maximal binding of MBP-VEGF1 to its target site, which in this experiment was approximately 2 nM.

The binding affinities determined for the VEGF proteins can be summarized as follows. VEGF1 showed the stronger DNA-binding affinity; in multiple EMSA analyses, the average apparent K_d was determined to be approximately 10 nM when bound to VEGF site 1. VEGF3a bound well to its target site but with a higher apparent K_d than VEGF1; the average K_d for VEGF3a was about 200 nM. In both cases

pure as estimated by staining of SDS polyacrylamide gels with Coomassie blue, and the product migrated at the predicted molecular weight of around 11 kDa (Figure 2).

The second method of producing ZFPs was to express them as fusions to the *E. coli* Maltose Binding Protein (MBP). N-terminal MBP fusions to the ZFPs were constructed by PCR amplification of the pET15b clones and insertion into the vector pMal-c2 under the control of the Tac promoter (New England Biolabs). The fusion allows simple purification and detection of the recombinant protein. It had been reported previously that zinc finger DNA-binding proteins can be expressed from this vector in soluble form to high levels in *E. coli* and can bind efficiently to the appropriate DNA target without refolding (Liu *et al. PNAS* 94:5525-5530 (1997)). Production of MBP-fused proteins was as described by the manufacturer (New England Biolabs). Transformants were grown in LB medium supplemented with glucose and ampicillin, and were induced with IPTG for 3 hrs at 37°C. The cells were lysed by French press, then exposed to an agarose-based amylose resin, which specifically binds to the MBP moiety, thus acting as an affinity resin for this protein. The MBP fusion protein was eluted with 10 mM maltose (Figure 2C) to release ZFP of >50% purity. In some cases, the proteins were further concentrated using a Centricon 30 filter unit (Amicon).

Partially purified unfused and MBP fusion ZFPs were tested by EMSA to assess binding to their target DNA sequences. The protein concentrations in the preparations were measured by Bradford assay (BioRad). Since SDS polyacrylamide gels demonstrated >50% homogeneity by either purification method, no adjustment was made for ZFP purity in the calculations. In addition, there could be significant amounts of inactive protein in the preparations. Therefore, the data generated by EMSAs below represent an underestimate of the true affinity of the proteins for their targets (i.e., overestimate of K_ds). Two separate preparations were made for each protein to help control for differences in ZFP activity.

The VEGF DNA target sites for the EMSA experiments were generated by embedding the 9-bp binding sites in 29-bp duplex oligonucleotides. The sequences of the recognition ("top") strand and their complements ("bottom") used in the assays are as follows:

VEGF site 1, top: 5'-CATGCATAGCGGGGAGGATCGCCATCGAT

VEGF site 1, bottom: 5'-ATCGATGGCGATCCTCCCCGCTATGCATG

TGCTTGCCCGGAGTGTCCGAAGCGCTTCATGCGTAGTGACCACCTGTCCCGTC
ACATCAAGACCCACCAGAATAAGAAGGGTGGATCC

VEGF1 translation:

VPIPGKKKQHICHIQGCGKVYGTTSNLRRHLRWHTGERPFMCTWSYCGKRFTRS
SNLQRHKRTHTGEKKFACPECPKRFMRSDHLSRHIKTHQNKKGGS

VEGF3a:

GGTACCCATACCTGGCAAGAAGAAGCAGCACATCTGCCACATCCAGGGCTGT
GGTAAAGTTTACGGCCAGTCCTCCGACCTGCAGCGTCACCTGCGCTGGCACAC
CGGCGAGAGGCCTTTCATGTGTACCTGGTCCTACTGTGGTAAACGCTTCACCC
GTTCGTCAAACCTACAGAGGCACAAGCGTACACACACCGGTGAGAAGAAATT
TGCTTGCCCGGAGTGTCCGAAGCGCTTCATGCGAAGTGACGAGCTGTCACGA
CATATCAAGACCCACCAGAACAAGAAGGGTGGATCC

VEGF3a translation:

VPIPGKKKQHICHIQGCGKVYGQSSDLQRHLRWHTGERPFMCTWSYCGKRFTRS
SNLQRHKRTHTGEKKFACPECPKRFMRSDELSRHIKTHQNKKGGS

The ability of the designed ZFPs to bind their target sites was verified by expressing and purifying recombinant protein from *E. coli* and performing electrophoretic mobility shift assays (EMSAs). The expression of ZFPs was carried out in two different systems. In the first, the DNA-binding peptides were expressed in *E. coli* by inserting them into the commercially available pET15b vector (Novagen). This vector contains a T7 promoter sequence to drive expression of the recombinant protein. The constructs were introduced into *E. coli* BL21/DE3 ($lacI^q$) cells, which contain an IPTG-inducible T7 polymerase. Cultures were supplemented with 50 μM $ZnCl_2$, were grown at 37°C to an OD at 600 nm of 0.5-0.6, and protein production was induced with IPTG for 2 hrs. ZFP expression was seen at very high levels, approximately 30% of total cellular protein (Figure 2). These proteins are referred to as "unfused" ZFPs.

Partially pure unfused ZFPs were produced as follows (adapted from Desjarlais & Berg, *Proteins: Structure, Function and Genetics* 12:101-104 (1992)). A frozen cell pellet was resuspended in 1/50th volume of 1 M NaCl, 25 mM Tris HCl (pH 8.0), 100 μM $ZnCl_2$, 5 mM DTT. The samples were boiled for 10 min. and centrifuged for 10 min. at ~3,000 x g. At this point the ZFP protein in the supernatant was > 50%

The selection of amino acids in the recognition helices of the two designed ZFPs, VEGF1 and VEGF3a, is summarized in Table 1.

Table 1

Amino acids chosen for recognition helices of VEGF-recognizing ZFPs

Position:	Finger 1				Finger 2				Finger 3			
Protein	-1	2	3	6	-1	2	3	6	-1	2	3	6
VEGF1	T	S	N	R	R	S	N	R	R	D	H	R
VEGF3A	Q	S	D	R	R	S	N	R	R	D	E	R

Coding sequences were constructed to express these peptides using a PCR-based assembly procedure that utilizes six overlapping oligonucleotides (Figure 1). Three oligonucleotides (oligos 1, 3, and 5 in Figure 1) corresponding to "universal" sequences that encode portions of the DNA-binding domain between the recognition helices. These oligonucleotides remain constant for any zinc finger construct. The other three "specific" oligonucleotides (oligos 2, 4, and 6 in Figure 1) were designed to encode the recognition helices. These oligonucleotides contained substitutions at positions -1, 2, 3 and 6 on the recognition helices to make them specific for each of the different DNA-binding domains. Codon bias was chosen to allow expression in both mammalian cells and *E. coli*.

The PCR synthesis was carried out in two steps. First, the double stranded DNA template was created by combining the six oligonucleotides (three universal, three specific) and using a four cycle PCR reaction with a low temperature (25°) annealing step. At this temperature, the six oligonucleotides join to form a DNA "scaffold." The gaps in the scaffold were filled in by a combination of Taq and Pfu polymerases. In the second phase of construction, the zinc finger template was amplified in thirty cycles by external primers that were designed to incorporate restriction sites for cloning into pUC19. Accuracy of clones for the VEGF ZFPs were verified by DNA sequencing. The DNA sequences of each of the two constructs are listed below.

VEGF1:

GGTACCCATACCTGGCAAGAAGAAGCAGCACATCTGCCACATCCAGGGCTGT
GGTAAAGTTTACGGCACAACCTCAAATCTGCGTCGTCACCTGCGCTGGCACAC
CGGCGAGAGGCCTTTCATGTGTACCTGGTCCTACTGTGGTAAACGCTTCACCC
GTTCGTCAAACCTGCAGCGTCACAAGCGTACCCACACCGGTGAGAAGAAATT

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

EXAMPLES

The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of noncritical parameters that could be changed or modified to yield essentially similar results.

Example I: Design and testing of ZFPs targeted to the human VEGF gene

This first Example demonstrates the construction of ZFPs designed to recognize DNA sequences contained in the promoter of the human vascular endothelial growth factor (VEGF) gene. VEGF is an approximately 46 kDa glycoprotein that is an endothelial cell-specific mitogen induced by hypoxia. VEGF has been implicated in angiogenesis associated with cancer, various retinopathies, and other serious diseases. The DNA target site chosen was a region surrounding the transcription initiation site of the gene. The two 9 base pair (bp) sites chosen are found within the sequence agcGGGGAGGATcGCGGAGGCTtgg, where the upper-case letters represent actual 9-bp targets. The protein targeting the upstream 9-bp target was denoted VEGF1, and the protein targeting the downstream 9-bp target was denoted VEGF3a. The major start site of transcription for VEGF is at the T at the 3' end of the first 9-bp target, which is underlined in the sequence above.

The human SP-1 transcription factor was used as a progenitor molecule for the construction of designed ZFPs. SP-1 has a three finger DNA-binding domain related to the well-studied murine Zif268 (Christy *et al., PNAS* 85:7857-7861 (1988)). Site-directed mutagenesis experiments using this domain have shown that the proposed "recognition rules" that operate in Zif268 can be used to adapt SP-1 to other target DNA sequences (Desjarlais & Berg, *PNAS* 91:11099-11103 (1994)). The SP-1 sequence used for construction of zinc finger clones corresponds to amino acids 533 to 624 in the SP-1 transcription factor.

Selection or induction during the course of development can in some cases trigger the substitution of one family member for another mutant member. This type of functional substitution may not be possible in the adult animal. A typical result of developmental compensation would be the lack of a phenotype in a knockout mouse when the ablation of that gene's function in an adult would otherwise cause a physiological change. This is a kind of false negative result that often confounds the interpretation of conventional knockout mouse models.

A few new methods have been developed to avoid embryonic lethality. These methods are typified by an approach using the cre recombinase and lox DNA recognition elements. The recognition elements are inserted into a gene of interest using homologous recombination (as described above) and the expression of the recombinase induced in adult mice post-development. This causes the deletion of a portion of the target gene and avoids developmental complications. The method is labor intensive and suffers form chimerism due to non-uniform induction of the recombinase.

The use of engineered ZFPs to manipulate gene expression can be restricted to adult animals using the small molecule regulated systems described in the previous section. Expression and/or function of a zinc finger-based repressor can be switched off during development and switched on at will in the adult animals. This approach relies on the addition of the ZFP expressing module only; homologous recombination is not required. Because the ZFP repressors are *trans* dominant, there is no concern about germline transmission or homozygosity. These issues dramatically affect the time and labor required to go from a poorly characterized gene candidate (a cDNA or EST clone) to a mouse model. This ability can be used to rapidly identify and/or validate gene targets for therapeutic intervention, generate novel model systems and permit the analysis of complex physiological phenomena (development, hematopoiesis, transformation, neural function etc.). Chimeric targeted mice can be derived according to Hogan *et al.*, *Manipulating the Mouse Embryo*: *A Laboratory Manual*, (1988); *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed., (1987); and Capecchi *et al.*, *Science* 244:1288 (1989.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

gene of interest or any transgene by placing the function and/or expression of a ZFP regulator under small molecule control.

Transgenic mice

A further application of the ZFP technology is manipulating gene expression in transgenic animals. As with cell lines, over-expression of an endogenous gene or the introduction of a heterologous gene to a transgenic animal, such as a transgenic mouse, is a fairly straightforward process. The ZFP technology is an improvement in these types of methods because one can circumvent the need for generating full-length cDNA clones of the gene under study.

Likewise, as with cell-based systems, conventional down-regulation of gene expression in transgenic animals is plagued by technical difficulties. Gene knockout by homologous recombination is the method most commonly applied currently. This method requires a relatively long genomic clone of the gene to be knocked out (ca. 10 kb). Typically, a selectable marker is inserted into an exon of the gene of interest to effect the gene disruption, and a second counter-selectable marker provided outside of the region of homology to select homologous versus non-homologous recombinants. This construct is transfected into embryonic stem cells and recombinants selected in culture. Recombinant stem cells are combined with very early stage embryos generating chimeric animals. If the chimerism extends to the germline homozygous knockout animals can be isolated by back-crossing. When the technology is successfully applied, knockout animals can be generated in approximately one year. Unfortunately two common issues often prevent the successful application of the knockout technology; embryonic lethality and developmental compensation. Embryonic lethality results when the gene to be knocked out plays an essential role in development. This can manifest itself as a lack of chimerism, lack of germline transmission or the inability to generate homozygous back crosses. Genes can play significantly different physiological roles during development versus in adult animals. Therefore, embryonic lethality is not considered a rationale for dismissing a gene target as a useful target for therapeutic intervention in adults. Embryonic lethality most often simply means that the gene of interest can not be easily studied in mouse models, using conventional methods.

Developmental compensation is the substitution of a related gene product for the gene product being knocked out. Genes often exist in extensive families.

difficult. Until now, simple methods for assigning function to differentially expressed genes have not kept pace with the ability to monitor differential gene expression.

Using conventional molecular approaches, over expression of a candidate gene can be accomplished by cloning a full-length cDNA, subcloning it into a mammalian expression vector and transfecting the recombinant vector into an appropriate host cell. This approach is straightforward but labor intensive, particularly when the initial candidate gene is represented by a simple expressed sequence tag (EST). Under expression of a candidate gene by "conventional" methods is yet more problematic. Antisense methods and methods that rely on targeted ribozymes are unreliable, succeeding for only a small fraction of the targets selected. Gene knockout by homologous recombination works fairly well in recombinogenic stem cells but very inefficiently in somatically derived cell lines. In either case large clones of syngeneic genomic DNA (on the order of 10 kb) should be isolated for recombination to work efficiently.

The ZFP technology can be used to rapidly analyze differential gene expression studies. Engineered ZFPs can be readily used to up or down-regulate any endogenous target gene. Very little sequence information is required to create a gene-specific DNA binding domain. This makes the ZFP technology ideal for analysis of long lists of poorly characterized differentially expressed genes. One can simply build a zinc finger-based DNA binding domain for each candidate gene, create chimeric up and down-regulating artificial transcription factors and test the consequence of up or down-regulation on the phenotype under study (transformation, response to a cytokine etc.) by switching the candidate genes on or off one at a time in a model system.

This specific example of using engineered ZFPs to add functional information to genomic data is merely illustrative. Any experimental situation that could benefit from the specific up or down-regulation of a gene or genes could benefit from the reliability and ease of use of engineered ZFPs.

Additionally, greater experimental control can be imparted by ZFPs than can be achieved by more conventional methods. This is because the production and/or function of an engineered ZFP can be placed under small molecule control. Examples of this approach are provided by the Tet-On system, the ecdysone-regulated system and a system incorporating a chimeric factor including a mutant progesterone receptor. These systems are all capable of indirectly imparting small molecule control on any endogenous

Agrobacterium tumefaciens-meditated transformation techniques are well described in the scientific literature (*see, e.g.*, Horsch *et al. Science* 233:496-498 (1984)); and Fraley *et al. PNAS* 80:4803 (1983)).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired ZFP-controlled phenotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the ZFP nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans *et al., Protoplasts Isolation and Culture, Handbook of Plant Cell Culture*, pp. 124-176 (1983); and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73 (1985). Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee *et al. Ann. Rev. of Plant Phys.* 38:467-486 (1987).

Functional genomics assays

ZFPs also have use for assays to determine the phenotypic consequences and function of gene expression. The recent advances in analytical techniques, coupled with focussed mass sequencing efforts have created the opportunity to identify and characterize many more molecular targets than were previously available. This new information about genes and their functions will speed along basic biological understanding and present many new targets for therapeutic intervention. In some cases analytical tools have not kept pace with the generation of new data. An example is provided by recent advances in the measurement of global differential gene expression. These methods, typified by gene expression microarrays, differential cDNA cloning frequencies, subtractive hybridization and differential display methods, can very rapidly identify genes that are up or down-regulated in different tissues or in response to specific stimuli. Increasingly, such methods are being used to explore biological processes such as, transformation, tumor progression, the inflammatory response, neurological disorders etc. One can now very easily generate long lists of differentially expressed genes that correlate with a given physiological phenomenon, but demonstrating a causative relationship between an individual differentially expressed gene and the phenomenon is

region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill.

Alternatively, the plant promoter may direct expression of the ZFP in a specific tissue or may be otherwise under more precise environmental or developmental control. Such promoters are referred to here as "inducible" promoters. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions or the presence of light.

Examples of promoters under developmental control include promoters that initiate transcription only in certain tissues, such as fruit, seeds, or flowers. For example, the use of a polygalacturonase promoter can direct expression of the ZFP in the fruit, a CHS-A (chalcone synthase A from petunia) promoter can direct expression of the ZFP in flower of a plant.

The vector comprising the ZFP sequences will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosluforon or Basta.

Such DNA constructs may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski *et al. EMBO J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm *et al. PNAS* 82:5824 (1985). Ballistic transformation techniques are described in Klein *et al. Nature* 327:70-73 (1987).

number of double bonds, or degree of saturation, determines the melting temperature, reactivity, cooking performance, and health attributes of the resulting oil.

The enzyme responsible for the conversion of oleic acid (18:1) into linoleic acid (18:2) (which is then the precursor for 18:3 formation) is Δ12-oleate desaturase, also referred to as omega-6 desaturase. A block at this step in the fatty acid desaturation pathway should result in the accumulation of oleic acid at the expense of polyunsaturates.

In one embodiment ZFPs are used to regulate expression of the FAD2-1 gene in soybeans. Two genes encoding microsomal Δ6 desaturases have been cloned recently from soybean, and are referred to as FAD2-1 and FAD2-2 (Heppard *et al., Plant Physiol.* 110:311-319 (1996)). FAD2-1 (delta 12 desaturase) appears to control the bulk of oleic acid desaturation in the soybean seed. ZFPs can thus be used to modulate gene expression of FAD2-1 in plants. Specifically, ZFPs can be used to inhibit expression of the FAD2-1 gene in soybean in order to increase the accumulation of oleic acid (18:1) in the oil seed. Moreover, ZFPs can be used to modulate expression of any other plant gene, such as delta-9 desaturase, delta-12 desaturases from other plants, delta-15 desaturase, acetyl-CoA carboxylase, acyl-ACP-thioesterase, ADP-glucose pyrophosphorylase, starch synthase, cellulose synthase, sucrose synthase, senescence-associated genes, heavy metal chelators, fatty acid hydroperoxide lyase, polygalacturonase, EPSP synthase, plant viral genes, plant fungal pathogen genes, and plant bacterial pathogen genes.

Recombinant DNA vectors suitable for transformation of plant cells are also used to deliver the ZFP of the invention to plant cells. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature (*see, e.g.*, Weising *et al. Ann. Rev. Genet.* 22:421-477 (1988)). A DNA sequence coding for the desired ZFP is combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the ZFP in the intended tissues of the transformed plant.

For example, a plant promoter fragment may be employed which will direct expression of the ZFP in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation

non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials. Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described.

Regulation of gene expression in plants

ZFPs can be used to engineer plants for traits such as increased disease resistance, modification of structural and storage polysaccharides, flavors, proteins, and fatty acids, fruit ripening, yield, color, nutritional characteristics, improved storage capability, and the like. In particular, the engineering of crop species for enhanced oil production, e.g., the modification of the fatty acids produced in oilseeds, is of interest.

Seed oils are composed primarily of triacylglycerols (TAGs), which are glycerol esters of fatty acids. Commercial production of these vegetable oils is accounted for primarily by six major oil crops (soybean, oil palm, rapeseed, sunflower, cotton seed, and peanut.) Vegetable oils are used predominantly (90%) for human consumption as margarine, shortening, salad oils, and frying oil. The remaining 10% is used for non-food applications such as lubricants, oleochemicals, biofuels, detergents, and other industrial applications.

The desired characteristics of the oil used in each of these applications varies widely, particularly in terms of the chain length and number of double bonds present in the fatty acids making up the TAGs. These properties are manipulated by the plant in order to control membrane fluidity and temperature sensitivity. The same properties can be controlled using ZFPs to produce oils with improved characteristics for food and industrial uses.

The primary fatty acids in the TAGs of oilseed crops are 16 to 18 carbons in length and contain 0 to 3 double bonds. Palmitic acid (16:0 [16 carbons: 0 double bonds]), oleic acid (18:1), linoleic acid (18:2), and linolenic acid (18:3) predominate. The

rheumatoid arthritis, psoriasis, HIV infection, sickle cell anemia, Alzheimer's disease, muscular dystrophy, neurodegenerative diseases, vascular disease, cystic fibrosis, stroke, and the like. Examples of microorganisms that can be inhibited by ZFP gene therapy include pathogenic bacteria, e.g., chlamydia, rickettsial bacteria, mycobacteria, staphylococci, streptococci, pneumococci, meningococci and conococci, klebsiella, proteus, serratia, pseudomonas, legionella, diphtheria, salmonella, bacilli, cholera, tetanus, botulism, anthrax, plague, leptospirosis, and Lyme disease bacteria; infectious fungus, e.g., *Aspergillus*, *Candida* species; protozoa such as sporozoa (e.g., *Plasmodia*), rhizopods (e.g., *Entamoeba*) and flagellates (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.);viral diseases, e.g., hepatitis (A, B, or C), herpes virus (e.g., VZV, HSV-1, HSV-6, HSV-II, CMV, and EBV), HIV, Ebola, adenovirus, influenza virus, flaviviruses, echovirus, rhinovirus, coxsackie virus, comovirus, respiratory syncytial virus, mumps virus, rotavirus, measles virus, rubella virus, parvovirus, vaccinia virus, HTLV virus, dengue virus, papillomavirus, poliovirus, rabies virus, and arboviral encephalitis virus, etc.

Administration of therapeutically effective amounts is by any of the routes normally used for introducing ZFP into ultimate contact with the tissue to be treated. The ZFPs are administered in any suitable manner, preferably with pharmaceutically acceptable carriers. Suitable methods of administering such modulators are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention (*see, e.g., Remington's Pharmaceutical Sciences*, 17th ed. 1985)).

The ZFPs, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

Formulations suitable for parenteral administration, such as, for example, by intravenous, intramuscular, intradermal, and subcutaneous routes, include aqueous and

competition for ZFP binding by other sites. This calculation also assumes that essentially all of the ZFP is localized to the nucleus. A value of 100x K_d is used to calculate approximately 99% binding of to the target site, and a value of 10x K_d is used to calculate approximately 90% binding of to the target site. For this example, K_d = 25 nM

ZFP + target site ↔ complex

i.e., DNA + protein ↔ DNA:protein complex

$$K_d = \frac{[\text{DNA}]\,[\text{protein}]}{[\text{DNA:protein complex}]}$$

When 50% of ZFP is bound, K_d = [protein]

So when [protein] = 25 nM and the nucleus volume is 10^{-12} L

[protein] = (25×10^{-9} moles/L) (10^{-12} L/nucleus) (6×10^{23} molecules/mole)

= 15,000 molecules/nucleus for 50% binding

When 99% target is bound; 100x K_d = [protein]

100x K_d = [protein] = 2.5 μM

(2.5×10^{-6} moles/L) (10^{-12}L/nucleus) (6×10^{23} molecules/mole)

= about 1,500,000 molecules per nucleus for 99% binding of target site.

The appropriate dose of an expression vector encoding a ZFP can also be calculated by taking into account the average rate of ZFP expression from the promoter and the average rate of ZFP degradation in the cell. Preferably, a weak promoter such as a wild-type or mutant HSV TK is used, as described above. The dose of ZFP in micrograms is calculated by taking into account the molecular weight of the particular ZFP being employed.

In determining the effective amount of the ZFP to be administered in the treatment or prophylaxis of disease, the physician evaluates circulating plasma levels of the ZFP or nucleic acid encoding the ZFP, potential ZFP toxicities, progression of the disease, and the production of anti-ZFP antibodies. Administration can be accomplished via single or divided doses.

Pharmaceutical compositions and administration

ZFPs and expression vectors encoding ZFPs can be administered directly to the patient for modulation of gene expression and for therapeutic or prophylactic applications, for example, cancer, ischemia, diabetic retinopathy, macular degeneration,

carcinoembryonic antigen (CEA). Sites of viral infection can be diagnosed using various viral antigens such as hepatitis B core and surface antigens (HBVc, HBVs) hepatitis C antigens, Epstein-Barr virus antigens, human immunodeficiency type-1 virus (HIV1) and papilloma virus antigens. Inflammation can be detected using molecules specifically recognized by surface molecules which are expressed at sites of inflammation such as integrins (e.g., VCAM-1), selectin receptors (e.g., ELAM-1) and the like.

Standard methods for coupling targeting agents to liposomes can be used. These methods generally involve incorporation into liposomes lipid components, e.g., phosphatidylethanolamine, which can be activated for attachment of targeting agents, or derivatized lipophilic compounds, such as lipid derivatized bleomycin. Antibody targeted liposomes can be constructed using, for instance, liposomes which incorporate protein A (*see* Renneisen *et al.*, *J. Biol. Chem.*, 265:16337-16342 (1990) and Leonetti *et al.*, *PNAS* 87:2448-2451 (1990).

Doses of ZFPs

For therapeutic applications of ZFPs, the dose administered to a patient, in the context of the present invention should be sufficient to effect a beneficial therapeutic response in the patient over time. In addition, particular dosage regimens can be useful for determining phenotypic changes in an experimental setting, e.g., in functional genomics studies, and in cell or animal models. The dose will be determined by the efficacy and K_d of the particular ZFP employed, the nuclear volume of the target cell, and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular patient.

The maximum therapeutically effective dosage of ZFP for approximately 99% binding to target sites is calculated to be in the range of less than about $1.5x10^5$ to $1.5x10^6$ copies of the specific ZFP molecule per cell. The number of ZFPs per cell for this level of binding is calculated as follows, using the volume of a HeLa cell nucleus (approximately 1000 μm^3 or 10^{-12} L; *Cell Biology*, (Altman & Katz, eds. (1976)). As the HeLa nucleus is relatively large, this dosage number is recalculated as needed using the volume of the target cell nucleus. This calculation also does not take into account

involves using an agent to induce a permeability change in the liposome vesicle. Liposome membranes can be constructed so that they become destabilized when the environment becomes acidic near the liposome membrane (*see, e.g., PNAS* 84:7851 (1987); *Biochemistry* 28:908 (1989)). When liposomes are endocytosed by a target cell, for example, they become destabilized and release their contents. This destabilization is termed fusogenesis. Dioleoylphosphatidylethanolamine (DOPE) is the basis of many "fusogenic" systems.

Such liposomes typically comprise a ZFP and a lipid component, e.g., a neutral and/or cationic lipid, optionally including a receptor-recognition molecule such as an antibody that binds to a predetermined cell surface receptor or ligand (e.g., an antigen). A variety of methods are available for preparing liposomes as described in, e.g., Szoka *et al., Ann. Rev. Biophys. Bioeng.* 9:467 (1980), U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,946,787, PCT Publication No. WO 91\17424, Deamer & Bangham, *Biochim. Biophys. Acta* 443:629-634 (1976); Fraley, *et al., PNAS* 76:3348-3352 (1979); Hope *et al., Biochim. Biophys. Acta* 812:55-65 (1985); Mayer *et al., Biochim. Biophys. Acta* 858:161-168 (1986); Williams *et al., PNAS* 85:242-246 (1988); *Liposomes* (Ostro (ed.), 1983, Chapter 1); Hope *et al., Chem. Phys. Lip.* 40:89 (1986); Gregoriadis, *Liposome Technology* (1984) and Lasic, *Liposomes: from Physics to Applications* (1993)). Suitable methods include, for example, sonication, extrusion, high pressure/homogenization, microfluidization, detergent dialysis, calcium-induced fusion of small liposome vesicles and ether-fusion methods, all of which are well known in the art.

In certain embodiments of the present invention, it is desirable to target the liposomes of the invention using targeting moieties that are specific to a particular cell type, tissue, and the like. Targeting of liposomes using a variety of targeting moieties (e.g., ligands, receptors, and monoclonal antibodies) has been previously described (*see, e.g.*, U.S. Patent Nos. 4,957,773 and 4,603,044).

Examples of targeting moieties include monoclonal antibodies specific to antigens associated with neoplasms, such as prostate cancer specific antigen and MAGE. Tumors can also be diagnosed by detecting gene products resulting from the activation or over-expression of oncogenes, such as ras or c-erbB2. In addition, many tumors express antigens normally expressed by fetal tissue, such as the alphafetoprotein (AFP) and

Toxin molecules also have the ability to transport polypeptides across cell membranes. Often, such molecules are composed of at least two parts (called "binary toxins"): a translocation or binding domain or polypeptide and a separate toxin domain or polypeptide. Typically, the translocation domain or polypeptide binds to a cellular receptor, and then the toxin is transported into the cell. Several bacterial toxins, including *Clostridium perfringens* iota toxin, diphtheria toxin (DT), *Pseudomonas* exotoxin A (PE), pertussis toxin (PT), *Bacillus anthracis* toxin, and pertussis adenylate cyclase (CYA), have been used in attempts to deliver peptides to the cell cytosol as internal or amino-terminal fusions (Arora *et al., J. Biol. Chem.*, 268:3334-3341 (1993); Perelle *et al., Infect. Immun.*, 61:5147-5156 (1993); Stenmark *et al., J. Cell Biol.* 113:1025-1032 (1991); Donnelly *et al., PNAS* 90:3530-3534 (1993); Carbonetti *et al., Abstr. Annu. Meet. Am. Soc. Microbiol.* 95:295 (1995); Sebo *et al., Infect. Immun.* 63:3851-3857 (1995); Klimpel *et al., PNAS U.S.A.* 89:10277-10281 (1992); and Novak *et al., J. Biol. Chem.* 267:17186-17193 1992)).

Such subsequences can be used to translocate ZFPs across a cell membrane. ZFPs can be conveniently fused to or derivatized with such sequences. Typically, the translocation sequence is provided as part of a fusion protein. Optionally, a linker can be used to link the ZFP and the translocation sequence. Any suitable linker can be used, e.g., a peptide linker.

The ZFP can also be introduced into an animal cell, preferably a mammalian cell, via a liposomes and liposome derivatives such as immunoliposomes. The term "liposome" refers to vesicles comprised of one or more concentrically ordered lipid bilayers, which encapsulate an aqueous phase. The aqueous phase typically contains the compound to be delivered to the cell, i.e., a ZFP.

The liposome fuses with the plasma membrane, thereby releasing the drug into the cytosol. Alternatively, the liposome is phagocytosed or taken up by the cell in a transport vesicle. Once in the endosome or phagosome, the liposome either degrades or fuses with the membrane of the transport vesicle and releases its contents.

In current methods of drug delivery via liposomes, the liposome ultimately becomes permeable and releases the encapsulated compound (in this case, a ZFP) at the target tissue or cell. For systemic or tissue specific delivery, this can be accomplished, for example, in a passive manner wherein the liposome bilayer degrades over time through the action of various agents in the body. Alternatively, active drug release

administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention, as described below (*see, e.g., Remington's Pharmaceutical Sciences*, 17th ed., 1989).

Delivery vehicles for ZFPs

An important factor in the administration of polypeptide compounds, such as the ZFPs, is ensuring that the polypeptide has the ability to traverse the plasma membrane of a cell, or the membrane of an intra-cellular compartment such as the nucleus. Cellular membranes are composed of lipid-protein bilayers that are freely permeable to small, nonionic lipophilic compounds and are inherently impermeable to polar compounds, macromolecules, and therapeutic or diagnostic agents. However, proteins and other compounds such as liposomes have been described, which have the ability to translocate polypeptides such as ZFPs across a cell membrane.

For example, "membrane translocation polypeptides" have amphiphilic or hydrophobic amino acid subsequences that have the ability to act as membrane-translocating carriers. In one embodiment, homeodomain proteins have the ability to translocate across cell membranes. The shortest internalizable peptide of a homeodomain protein, Antennapedia, was found to be the third helix of the protein, from amino acid position 43 to 58 (*see, e.g.*, Prochiantz, *Current Opinion in Neurobiology* 6:629-634 (1996)). Another subsequence, the h (hydrophobic) domain of signal peptides, was found to have similar cell membrane translocation characteristics (*see, e.g.*, Lin *et al., J. Biol. Chem.* 270:1 4255-14258 (1995)).

Examples of peptide sequences which can be linked to a ZFP of the invention, for facilitating uptake of ZFP into cells, include, but are not limited to: an 11 animo acid peptide of the tat protein of HIV; a 20 residue peptide sequence which corresponds to amino acids 84-103 of the p16 protein (*see* Fahraeus *et al., Current Biology* 6:84 (1996)); the third helix of the 60-amino acid long homeodomain of Antennapedia (Derossi *et al., J. Biol. Chem.* 269:10444 (1994)); the h region of a signal peptide such as the Kaposi fibroblast growth factor (K-FGF) h region (Lin *et al., supra*); or the VP22 translocation domain from HSV (Elliot & O'Hare, *Cell* 88:223-233 (1997)). Other suitable chemical moieties that provide enhanced cellular uptake may also be chemically linked to ZFPs.

from an individual patient (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

Ex vivo cell transfection for diagnostics, research, or for gene therapy (e.g., via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In a preferred embodiment, cells are isolated from the subject organism, transfected with a ZFP nucleic acid (gene or cDNA), and re-infused back into the subject organism (e.g., patient). Various cell types suitable for *ex vivo* transfection are well known to those of skill in the art (*see, e.g.*, Freshney *et al., Culture of Animal Cells, A Manual of Basic Technique* (3rd ed. 1994)) and the references cited therein for a discussion of how to isolate and culture cells from patients).

In one embodiment, stem cells are used in *ex vivo* procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types *in vitro*, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ cells *in vitro* into clinically important immune cell types using cytokines such a GM-CSF, IFN-γ and TNF-α are known (*see* Inaba *et al., J. Exp. Med.* 176:1693-1702 (1992)).

Stem cells are isolated for transduction and differentiation using known methods. For example, stem cells are isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panB cells), GR-1 (granulocytes), and Iad (differentiated antigen presenting cells) (*see* Inaba *et al., J. Exp. Med.* 176:1693-1702 (1992)).

Vectors (e.g., retroviruses, adenoviruses, liposomes, etc.) containing therapeutic ZFP nucleic acids can be also administered directly to the organism for transduction of cells *in vivo*. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to

cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are supplied in *trans* by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. A viral vector is typically modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the viruses outer surface. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han *et al.*, *PNAS* 92:9747-9751 (1995), reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, filamentous phage can be engineered to display antibody fragments (e.g., FAB or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

Gene therapy vectors can be delivered *in vivo* by administration to an individual patient, typically by systemic administration (e.g., intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted

system. All of these viral vectors utilize approaches that involve complementation of defective vectors by genes inserted into helper cell lines to generate the transducing agent.

pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials (Dunbar *et al., Blood* 85:3048-305 (1995); Kohn *et al., Nat. Med.* 1:1017-102 (1995); Malech *et al., PNAS* 94:22 12133-12138 (1997)). PA317/pLASN was the first therapeutic vector used in a gene therapy trial. (Blaese *et al., Science* 270:475-480 (1995)). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors. (Ellem *et al., Immunol Immunother.* 44(1):10-20 (1997); Dranoff *et al., Hum. Gene Ther.* 1:111-2 (1997).

Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. All vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. (Wagner *et al., Lancet* 351:9117 1702-3 (1998), Kearns *et al., Gene Ther.* 9:748-55 (1996)).

Replication-deficient recombinant adenoviral vectors (Ad) are predominantly used for colon cancer gene therapy, because they can be produced at high titer and they readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the Ad E1a, E1b, and E3 genes; subsequently the replication defector vector is propagated in human 293 cells that supply deleted gene function in trans. Ad vectors can transduce multiply types of tissues *in vivo*, including nondividing, differentiated cells such as those found in the liver, kidney and muscle system tissues. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection (Sterman *et al., Hum. Gene Ther.* 7:1083-9 (1998)). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker *et al., Infection* 24:1 5-10 (1996); Sterman *et al., Hum. Gene Ther.* 9:7 1083-1089 (1998); Welsh *et al., Hum. Gene Ther.* 2:205-18 (1995); Alvarez *et al., Hum. Gene Ther.* 5:597-613 (1997); Topf *et al., Gene Ther.* 5:507-513 (1998); Sterman *et al., Hum. Gene Ther.* 7:1083-1089 (1998).

Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2

the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vector that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of *cis*-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum *cis*-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (*see, e.g.*, Buchscher *et al.*, *J. Virol.* 66:2731-2739 (1992); Johann *et al.*, *J. Virol.* 66:1635-1640 (1992); Sommerfelt *et al.*, *Virol.* 176:58-59 (1990); Wilson *et al.*, *J. Virol.* 63:2374-2378 (1989); Miller *et al.*, *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700).

In applications where transient expression of the ZFP is preferred, adenoviral based systems are typically used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, e.g., in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (*see, e.g.*, West *et al.*, *Virology* 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin *et al.*, *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, *et al.*, *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulski *et al.*, *J. Virol.* 63:03822-3828 (1989).

In particular, at least six viral vector approaches are currently available for gene transfer in clinical trials, with retroviral vectors by far the most frequently used

(1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada *et al.*, in *Current Topics in Microbiology and Immunology* Doerfler and Böhm (eds) (1995); and Yu *et al.*, *Gene Therapy* 1:13-26 (1994).

Methods of non-viral delivery of nucleic acids encoding engineered ZFPs include lipofection, microinjection, ballistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., US 5,049,386, US 4,946,787; and US 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024. Delivery can be to cells (*ex vivo* administration) or target tissues (*in vivo* administration).

The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (*see, e.g.*, Crystal, *Science* 270:404-410 (1995); Blaese *et al.*, *Cancer Gene Ther.* 2:291-297 (1995); Behr *et al.*, *Bioconjugate Chem.* 5:382-389 (1994); Remy *et al.*, *Bioconjugate Chem.* 5:647-654 (1994); Gao *et al.*, *Gene Therapy* 2:710-722 (1995); Ahmad *et al.*, *Cancer Res.* 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding engineered ZFP take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (*in vivo*) or they can be used to treat cells *in vitro* and the modified cells are administered to patients (*ex vivo*). Conventional viral based systems for the delivery of ZFPs could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Viral vectors are currently the most efficient and versatile method of gene transfer in target cells and tissues. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of

Another example of a preferred assay format useful for monitoring ZFP regulation of endogenous gene expression is performed *in vivo*. This assay is particularly useful for examining ZFPs that inhibit expression of tumor promoting genes, genes involved in tumor support, such as neovascularization (e.g., VEGF), or that activate tumor suppressor genes such as p53. In this assay, cultured tumor cells expressing the ZFP of choice are injected subcutaneously into an immune compromised mouse such as an athymic mouse, an irradiated mouse, or a SCID mouse. After a suitable length of time, preferably 4-8 weeks, tumor growth is measured, e.g., by volume or by its two largest dimensions, and compared to the control. Tumors that have statistically significant reduction (using, e.g., Student's T test) are said to have inhibited growth. Alternatively, the extent of tumor neovascularization can also be measured. Immunoassays using endothelial cell specific antibodies are used to stain for vascularization of the tumor and the number of vessels in the tumor. Tumors that have a statistically significant reduction in the number of vessels (using, e.g., Student's T test) are said to have inhibited neovascularization.

Transgenic and non-transgenic animals are also used as a preferred embodiment for examining regulation of endogenous gene expression *in vivo*. Transgenic animals typically express the ZFP of choice. Alternatively, animals that transiently express the ZFP of choice, or to which the ZFP has been administered in a delivery vehicle, can be used. Regulation of endogenous gene expression is tested using any one of the assays described herein.

Nucleic acids encoding ZFPs and gene therapy

Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding engineered ZFP in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding ZFPs to cells *in vitro*. Preferably, the nucleic acids encoding ZFPs are administered for *in vivo* or *ex vivo* gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, *Science* 256:808-813 (1992); Nabel & Felgner, *TIBTECH* 11:211-217 (1993); Mitani & Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175

Modulation of gene expression is tested using one of the *in vitro* or *in vivo* assays described herein. Samples or assays are treated with a ZFP and compared to control samples without the test compound, to examine the extent of modulation. As described above, for regulation of endogenous gene expression, the ZFP typically has a K_d of 200 nM or less, more preferably 100 nM or less, more preferably 50 nM, most preferably 25 nM or less.

The effects of the ZFPs can be measured by examining any of the parameters described above. Any suitable gene expression, phenotypic, or physiological change can be used to assess the influence of a ZFP. When the functional consequences are determined using intact cells or animals, one can also measure a variety of effects such as tumor growth, neovascularization, hormone release, transcriptional changes to both known and uncharacterized genetic markers (e.g., northern blots or oligonucleotide array studies), changes in cell metabolism such as cell growth or pH changes, and changes in intracellular second messengers such as cGMP.

Preferred assays for ZFP regulation of endogenous gene expression can be performed *in vitro*. In one preferred *in vitro* assay format, ZFP regulation of endogenous gene expression in cultured cells is measured by examining protein production using an ELISA assay (*see* Examples VI and VII). The test sample is compared to control cells treated with an empty vector or an unrelated ZFP that is targeted to another gene.

In another embodiment, ZFP regulation of endogenous gene expression is determined *in vitro* by measuring the level of target gene mRNA expression. The level of gene expression is measured using amplification, e.g., using PCR, LCR, or hybridization assays, e.g., northern hybridization, RNase protection, dot blotting. RNase protection is used in one embodiment (*see* Example VIII and Figure 10). The level of protein or mRNA is detected using directly or indirectly labeled detection agents, e.g., fluorescently or radioactively labeled nucleic acids, radioactively or enzymatically labeled antibodies, and the like, as described herein.

Alternatively, a reporter gene system can be devised using the target gene promoter operably linked to a reporter gene such as luciferase, green fluorescent protein, CAT, or β-gal. The reporter construct is typically co-transfected into a cultured cell. After treatment with the ZFP of choice, the amount of reporter gene transcription, translation, or activity is measured according to standard techniques known to those of skill in the art.

to standard techniques (*see, e.g.*, Morrison, *J. Bact.* 132:349-351 (1977); Clark-Curtiss & Curtiss, *Methods in Enzymology* 101:347-362 (Wu *et al.*, eds, 1983).

Any of the well known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see, e.g.*, Sambrook *et al.*, *supra*). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the protein of choice.

Assays for determining regulation of gene expression by ZFPs

A variety of assays can be used to determine the level of gene expression regulation by ZFPs. The activity of a particular ZFP can be assessed using a variety of *in vitro* and *in vivo* assays, by measuring, e.g., protein or mRNA levels, product levels, enzyme activity, tumor growth; transcriptional activation or repression of a reporter gene; second messenger levels (e.g., cGMP, cAMP, IP3, DAG, Ca^{2+}); cytokine and hormone production levels; and neovascularization, using, e.g., immunoassays (e.g., ELISA and immunohistochemical assays with antibodies), hybridization assays (e.g., RNase protection, northerns, *in situ* hybridization, oligonucleotide array studies), colorimetric assays, amplification assays, enzyme activity assays, tumor growth assays, phenotypic assays, and the like.

ZFPs are typically first tested for activity *in vitro* using cultured cells, e.g., 293 cells, CHO cells, VERO cells, BHK cells, HeLa cells, COS cells, and the like. Preferably, human cells are used. The ZFP is often first tested using a transient expression system with a reporter gene, and then regulation of the target endogenous gene is tested in cells and in animals, both *in vivo* and *ex vivo*. The ZFP can be recombinantly expressed in a cell, recombinantly expressed in cells transplanted into an animal, or recombinantly expressed in a transgenic animal, as well as administered as a protein to an animal or cell using delivery vehicles described below. The cells can be immobilized, be in solution, be injected into an animal, or be naturally occurring in a transgenic or non-transgenic animal.

The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the ZFP, e.g., expression in plants, animals, bacteria, fungus, protozoa etc. (see expression vectors described below and in the Example section). Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and commercially available fusion expression systems such as GST and LacZ. A preferred fusion protein is the maltose binding protein, "MBP." Such fusion proteins are used for purification of the ZFP. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, for monitoring expression, and for monitoring cellular and subcellular localization, e.g., c-myc or FLAG.

Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High yield expression systems are also suitable, such as using a baculovirus vector in insect cells, with a ZFP encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of recombinant sequences.

Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of protein, which are then purified using standard techniques (*see, e.g.,* Colley *et al., J. Biol. Chem.* 264:17619-17622 (1989); *Guide to Protein Purification, in Methods in Enzymology*, vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according

plant cell, animal cell, preferably a mammalian cell or a human cell, fungal cell, bacterial cell, or protozoal cell.

To obtain expression of a cloned gene or nucleic acid, a ZFP is typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, e.g., in Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994). Bacterial expression systems for expressing the ZFP are available in, e.g., *E. coli*, *Bacillus sp.*, and *Salmonella* (Palva *et al.*, *Gene* 22:229-235 (1983)). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available.

The promoter used to direct expression of a ZFP nucleic acid depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification of ZFP. In contrast, when a ZFP is administered *in vivo* for gene regulation, either a constitutive or an inducible promoter is used, depending on the particular use of the ZFP. In addition, a preferred promoter for administration of a ZFP can be a weak promoter, such as HSV TK or a promoter having similar activity. The promoter typically can also include elements that are responsive to transactivation, e.g., hypoxia response elements, Gal4 response elements, lac repressor response element, and small molecule control systems such as tet-regulated systems and the RU-486 system (*see, e.g.*, Gossen & Bujard, *PNAS* 89:5547 (1992); Oligino *et al.*, *Gene Ther.* 5:491-496 (1998); Wang *et al.*, *Gene Ther.* 4:432-441 (1997); Neering *et al.*, *Blood* 88:1147-1155 (1996); and Rendahl *et al.*, *Nat. Biotechnol.* 16:757-761 (1998)).

In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter operably linked, e.g., to the nucleic acid sequence encoding the ZFP, and signals required, e.g., for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding sites, or translation termination. Additional elements of the cassette may include, e.g., enhancers, and heterologous spliced intronic signals.

polypeptide sequences, such as polyglycine sequences of between about 5 and 200 amino acids. Preferred linkers are typically flexible amino acid subsequences which are synthesized as part of a recombinant fusion protein. For example, in one embodiment, the linker DGGGS is used to link two ZFPs. In another embodiment, the flexible linker linking two ZFPs is an amino acid subsequence comprising the sequence TGEKP (*see, e.g.*, Liu *et al.*, *PNAS* 5525-5530 (1997)). In another embodiment, the linker LRQKDGERP is used to link two ZFPs. In another embodiment, the following linkers are used to link two ZFPs: GGRR (Pomerantz *et al.* 1995, *supra*), $(G4S)_n$ (Kim *et al.*, *PNAS* 93, 1156-1160 (1996); and GGRRGGGS; LRQRDGERP; LRQKDGGGSERP; LRQKD$(G3S)_2$ ERP. Alternatively, flexible linkers can be rationally designed using computer program capable of modeling both DNA-binding sites and the peptides themselves (Desjarlais & Berg, *PNAS* 90:2256-2260 (1993), *PNAS* 91:11099-11103 (1994) or by phage display methods.

In other embodiments, a chemical linker is used to connect synthetically or recombinantly produced domain sequences. Such flexible linkers are known to persons of skill in the art. For example, poly(ethylene glycol) linkers are available from Shearwater Polymers, Inc. Huntsville, Alabama. These linkers optionally have amide linkages, sulfhydryl linkages, or heterofunctional linkages. In addition to covalent linkage of ZFPs to regulatory domains, non-covalent methods can be used to produce molecules with ZFPs associated with regulatory domains.

In addition to regulatory domains, often the ZFP is expressed as a fusion protein such as maltose binding protein ("MBP"), glutathione S transferase (GST), hexahistidine, c-myc, and the FLAG epitope, for ease of purification, monitoring expression, or monitoring cellular and subcellular localization.

Expression vectors for nucleic acids encoding ZFP

The nucleic acid encoding the ZFP of choice is typically cloned into intermediate vectors for transformation into prokaryotic or eukaryotic cells for replication and/or expression, e.g., for determination of K_d. Intermediate vectors are typically prokaryote vectors, e.g., plasmids, or shuttle vectors, or insect vectors, for storage or manipulation of the nucleic acid encoding ZFP or production of protein. The nucleic acid encoding a ZFP is also typically cloned into an expression vector, for administration to a

example, Vos, *Curr. Opin. Cell Biol.* 4:385-95 (1992); Sancar, *Ann. Rev. Genet.* 29:69-105 (1995); Lehmann, *Genet. Eng.* 17:1-19 (1995); and Wood, *Ann. Rev. Biochem.* 65:135-67 (1996). DNA rearrangement enzymes and their associated factors and modifiers can also be used as regulatory domains (*see, e.g.,* Gangloff *et al., Experientia* 50:261-9 (1994); Sadowski, *FASEB J.* 7:760-7 (1993)).

Similarly, regulatory domains can be derived from DNA modifying enzymes (e.g., DNA methyltransferases, topoisomerases, helicases, ligases, kinases, phosphatases, polymerases) and their associated factors and modifiers. Helicases are reviewed in Matson *et al., Bioessays,* 16:13-22 (1994), and methyltransferases are described in Cheng, *Curr. Opin. Struct. Biol.* 5:4-10 (1995). Chromatin associated proteins and their modifiers (e.g., kinases, acetylases and deacetylases), such as histone deacetylase (Wolffe, *Science* 272:371-2 (1996)) are also useful as domains for addition to the ZFP of choice. In one preferred embodiment, the regulatory domain is a DNA methyl transferase that acts as a transcriptional repressor (*see, e.g.,* Van den Wyngaert *et al., FEBS Lett.* 426:283-289 (1998); Flynn *et al., J. Mol. Biol.* 279:101-116 (1998); Okano *et al., Nucleic Acids Res.* 26:2536-2540 (1998); and Zardo & Caiafa, *J. Biol. Chem.* 273:16517-16520 (1998)). In another preferred embodiment, endonucleases such as Fok1 are used as transcriptional repressors, which act via gene cleavage (*see, e.g.,* WO95/09233; and PCT/US94/01201).

Factors that control chromatin and DNA structure, movement and localization and their associated factors and modifiers; factors derived from microbes (e.g., prokaryotes, eukaryotes and virus) and factors that associate with or modify them can also be used to obtain chimeric proteins. In one embodiment, recombinases and integrases are used as regulatory domains. In one embodiment, histone acetyltransferase is used as a transcriptional activator (*see, e.g.,* Jin & Scotto, *Mol. Cell. Biol.* 18:4377-4384 (1998); Wolffe, *Science* 272:371-372 (1996); Taunton *et al., Science* 272:408-411 (1996); and Hassig *et al., PNAS* 95:3519-3524 (1998)). In another embodiment, histone deacetylase is used as a transcriptional repressor (*see, e.g.,* Jin & Scotto, *Mol. Cell. Biol.* 18:4377-4384 (1998); Syntichaki & Thireos, *J. Biol. Chem.* 273:24414-24419 (1998); Sakaguchi *et al., Genes Dev.* 12:2831-2841 (1998); and Martinez *et al., J. Biol. Chem.* 273:23781-23785 (1998)).

Linker domains between polypeptide domains, e.g., between two ZFPs or between a ZFP and a regulatory domain, can be included. Such linkers are typically

In one embodiment, the HSV VP16 activation domain is used as a transcriptional activator (*see, e.g.*, Hagmann *et al., J. Virol.* 71:5952-5962 (1997)). Other preferred transcription factors that could supply activation domains include the VP64 activation domain (Seipel *et al., EMBO J.* 11:4961-4968 (1996)); nuclear hormone receptors (*see, e.g.*, Torchia *et al., Curr. Opin. Cell. Biol.* 10:373-383 (1998)); the p65 subunit of nuclear factor kappa B (Bitko & Barik, *J. Virol.* 72:5610-5618 (1998) and Doyle & Hunt, *Neuroreport* 8:2937-2942 (1997)); and EGR-1 (early growth response gene product-1; Yan *et al., PNAS* 95:8298-8303 (1998); and Liu *et al., Cancer Gene Ther.* 5:3-28 (1998)).

Kinases, phosphatases, and other proteins that modify polypeptides involved in gene regulation are also useful as regulatory domains for ZFPs. Such modifiers are often involved in switching on or off transcription mediated by, for example, hormones. Kinases involved in transcription regulation are reviewed in Davis, *Mol. Reprod. Dev.* 42:459-67 (1995), Jackson *et al., Adv. Second Messenger Phosphoprotein Res.* 28:279-86 (1993), and Boulikas, *Crit. Rev. Eukaryot. Gene Expr.* 5:1-77 (1995), while phosphatases are reviewed in, for example, Schonthal & Semin, *Cancer Biol.* 6:239-48 (1995). Nuclear tyrosine kinases are described in Wang, *Trends Biochem. Sci.* 19:373-6 (1994).

As described, useful domains can also be obtained from the gene products of oncogenes (e.g., myc, jun, fos, myb, max, mad, rel, ets, bcl, mos family members) and their associated factors and modifiers. Oncogenes are described in, for example, Cooper, *Oncogenes*, 2nd ed., The Jones and Bartlett Series in Biology, Boston, MA, Jones and Bartlett Publishers, 1995. The ets transcription factors are reviewed in Waslylk *et al.*, *Eur. J. Biochem.* 211:7-18 (1993) and Crepieux *et al., Crit. Rev. Oncog.* 5:615-38 (1994). Myc oncogenes are reviewed in, for example, Ryan *et al., Biochem. J.* 314:713-21 (1996). The jun and fos transcription factors are described in, for example, *The Fos and Jun Families of Transcription Factors*, Angel & Herrlich, eds. (1994). The max oncogene is reviewed in Hurlin *et al.*, Cold Spring Harb. Symp. Quant. Biol. 59:109-16. The myb gene family is reviewed in Kanei-Ishii *et al., Curr. Top. Microbiol. Immunol.* 211:89-98 (1996). The mos family is reviewed in Yew *et al., Curr. Opin. Genet. Dev.* 3:19-25 (1993).

ZFPs can include regulatory domains obtained from DNA repair enzymes and their associated factors and modifiers. DNA repair systems are reviewed in, for

factors are known (*see, e.g., Science* 269:630 (1995)). Nuclear hormone receptor transcription factors are described in, for example, Rosen *et al., J. Med. Chem.* 38:4855-74 (1995). The C/EBP family of transcription factors are reviewed in Wedel *et al., Immunobiology* 193:171-85 (1995). Coactivators and co-repressors that mediate transcription regulation by nuclear hormone receptors are reviewed in, for example, Meier, *Eur. J. Endocrinol.* 134(2):158-9 (1996); Kaiser *et al., Trends Biochem. Sci.* 21:342-5 (1996); and Utley *et al., Nature* 394:498-502 (1998)). GATA transcription factors, which are involved in regulation of hematopoiesis, are described in, for example, Simon, *Nat. Genet.* 11:9-11 (1995); Weiss *et al., Exp. Hematol.* 23:99-107. TATA box binding protein (TBP) and its associated TAF polypeptides (which include TAF30, TAF55, TAF80, TAF110, TAF150, and TAF250) are described in Goodrich & Tjian, *Curr. Opin. Cell Biol.* 6:403-9 (1994) and Hurley, *Curr. Opin. Struct. Biol.* 6:69-75 (1996). The STAT family of transcription factors are reviewed in, for example, Barahmand-Pour *et al., Curr. Top. Microbiol. Immunol.* 211:121-8 (1996). Transcription factors involved in disease are reviewed in Aso *et al., J. Clin. Invest.* 97:1561-9 (1996).

In one embodiment, the KRAB repression domain from the human KOX-1 protein is used as a transcriptional repressor (Thiesen *et al., New Biologist* 2:363-374 (1990); Margolin *et al., PNAS* 91:4509-4513 (1994); Pengue *et al., Nucl. Acids Res.* 22:2908-2914 (1994); Witzgall *et al., PNAS* 91:4514-4518 (1994); *see also* Example III)). In another embodiment, KAP-1, a KRAB co-repressor, is used with KRAB (Friedman *et al., Genes Dev.* 10:2067-2078 (1996)). Alternatively, KAP-1 can be used alone with a ZFP. Other preferred transcription factors and transcription factor domains that act as transcriptional repressors include MAD (*see, e.g.*, Sommer *et al., J. Biol. Chem.* 273:6632-6642 (1998); Gupta *et al., Oncogene* 16:1149-1159 (1998); Queva *et al., Oncogene* 16:967-977 (1998); Larsson *et al., Oncogene* 15:737-748 (1997); Laherty *et al., Cell* 89:349-356 (1997); and Cultraro *et al., Mol Cell. Biol.* 17:2353-2359 (19977)); FKHR (forkhead in rhapdosarcoma gene; Ginsberg *et al., Cancer Res.* 15:3542-3546 (1998); Epstein *et al., Mol. Cell. Biol.* 18:4118-4130 (1998)); EGR-1 (early growth response gene product-1; Yan *et al., PNAS* 95:8298-8303 (1998); and Liu *et al., Cancer Gene Ther.* 5:3-28 (1998)); the ets2 repressor factor repressor domain (ERD; Sgouras *et al., EMBO J.* 14:4781-4793 ((19095)); and the MAD smSIN3 interaction domain (SID; Ayer *et al., Mol. Cell. Biol.* 16:5772-5781 (1996)).

then repeated as many times as is necessary to sufficiently enrich the phage pool for tight binders such that these may be identified using sequencing and/or screening methods.

Regulatory domains

The ZFPs of the invention can optionally be associated with regulatory domains for modulation of gene expression. The ZFP can be covalently or non-covalently associated with one or more regulatory domains, alternatively two or more regulatory domains, with the two or more domains being two copies of the same domain, or two different domains. The regulatory domains can be covalently linked to the ZFP, e.g., via an amino acid linker, as part of a fusion protein. The ZFPs can also be associated with a regulatory domain via a non-covalent dimerization domain, e.g., a leucine zipper, a STAT protein N terminal domain, or an FK506 binding protein (*see, e.g.*, O'Shea, *Science* 254: 539 (1991), Barahmand-Pour *et al.*, *Curr. Top. Microbiol. Immunol.* 211:121-128 (1996); Klemm *et al.*, *Annu. Rev. Immunol.* 16:569-592 (1998); Klemm *et al.*, *Annu. Rev. Immunol.* 16:569-592 (1998); Ho *et al.*, *Nature* 382:822-826 (1996); and Pomeranz *et al.*, *Biochem.* 37:965 (1998)). The regulatory domain can be associated with the ZFP at any suitable position, including the C- or N-terminus of the ZFP.

Common regulatory domains for addition to the ZFP include, e.g., effector domains from transcription factors (activators, repressors, co-activators, co-repressors), silencers, nuclear hormone receptors, oncogene transcription factors (e.g., myc, jun, fos, myb, max, mad, rel, ets, bcl, mos family members etc.); DNA repair enzymes and their associated factors and modifiers; DNA rearrangement enzymes and their associated factors and modifiers; chromatin associated proteins and their modifiers (e.g., kinases, acetylases and deacetylases); and DNA modifying enzymes (e.g., methyltransferases, topoisomerases, helicases, ligases, kinases, phosphatases, polymerases, endonucleases) and their associated factors and modifiers.

Transcription factor polypeptides from which one can obtain a regulatory domain include those that are involved in regulated and basal transcription. Such polypeptides include transcription factors, their effector domains, coactivators, silencers, nuclear hormone receptors (*see, e.g.*, Goodrich *et al.*, *Cell* 84:825-30 (1996) for a review of proteins and nucleic acid elements involved in transcription; transcription factors in general are reviewed in Barnes & Adcock, *Clin. Exp. Allergy* 25 Suppl. 2:46-9 (1995) and Roeder, *Methods Enzymol.* 273:165-71 (1996)). Databases dedicated to transcription

Similar assays can also include determining active fractions in the protein preparations. Active fractions are determined by stoichiometric gel shifts where proteins are titrated against a high concentration of target DNA. Titrations are done at 100, 50, and 25% of target (usually at micromolar levels).

In another embodiment, phage display libraries can be used to select ZFPs with high affinity to the selected target site. This method differs fundamentally from direct design in that it involves the generation of diverse libraries of mutagenized ZFPs, followed by the isolation of proteins with desired DNA-binding properties using affinity selection methods. To use this method, the experimenter typically proceeds as follows.

First, a gene for a ZFP is mutagenized to introduce diversity into regions important for binding specificity and/or affinity. In a typical application, this is accomplished via randomization of a single finger at positions –1, +2, +3, and +6, and perhaps accessory positions such as +1, +5, +8, or +10.

Next, the mutagenized gene is cloned into a phage or phagemid vector as a fusion with, e.g., gene III of filamentous phage, which encodes the coat protein pIII. The zinc finger gene is inserted between segments of gene III encoding the membrane export signal peptide and the remainder of pIII, so that the ZFP is expressed as an amino-terminal fusion with pIII in the mature, processed protein. When using phagemid vectors, the mutagenized zinc finger gene may also be fused to a truncated version of gene III encoding, minimally, the C-terminal region required for assembly of pIII into the phage particle.

The resultant vector library is transformed into *E. coli* and used to produce filamentous phage which express variant ZFPs on their surface as fusions with the coat protein pIII (if a phagemid vector is used, then the this step requires superinfection with helper phage). The phage library is then incubated with target DNA site, and affinity selection methods are used to isolate phage which bind target with high affinity from bulk phage. Typically, the DNA target is immobilized on a solid support, which is then washed under conditions sufficient to remove all but the tightest binding phage. After washing, any phage remaining on the support are recovered via elution under conditions which totally disrupt zinc finger-DNA binding.

Recovered phage are used to infect fresh *E. coli*, which is then amplified and used to produce a new batch of phage particles. The binding and recovery steps are

protein can be obtained by induction with IPTG since the MBP-ZFP fusion in the pMal-c2 expression plasmid is under the control of the IPTG inducible tac promoter (NEB). Bacteria containing the MBP-ZFP fusion plasmids are inoculated in to 2xYT medium containing 10μM $ZnCl_2$, 0.02% glucose, plus 50 μg/ml ampicillin and shaken at 37°C. At mid-exponential growth IPTG is added to 0.3 mM and the cultures are allowed to shake. After 3 hours the bacteria are harvested by centrifugation, disrupted by sonication, and then insoluble material is removed by centrifugation. The MBP-ZFP proteins are captured on an amylose-bound resin, washed extensively with buffer containing 20 mM Tris-HCl (pH 7.5), 200 mM NaCl, 5 mM DTT and 50 μM $ZnCl_2$, then eluted with maltose in essentially the same buffer (purification is based on a standard protocol from NEB). Purified proteins are quantitated and stored for biochemical analysis.

The biochemical properties of the purified proteins, e.g., K_d, can be characterized by any suitable assay. In one embodiment, K_d is characterized via electrophoretic mobility shift assays ("EMSA") (Buratowski & Chodosh, *in Current Protocols in Molecular Biology* pp. 12.2.1-12.2.7 (Ausubel ed., 1996); *see also* U.S. Patent No. 5,789,538, USSN 09/229,007, filed January 12, 1999, herein incorporated by reference, and Example I). Affinity is measured by titrating purified protein against a low fixed amount of labeled double-stranded oligonucleotide target. The target comprises the natural binding site sequence (9 or 18 bp) flanked by the 3 bp found in the natural sequence. External to the binding site plus flanking sequence is a constant sequence. The annealed oligonucleotide targets possess a 1 bp 5' overhang which allows for efficient labeling of the target with T4 phage polynucleotide kinase. For the assay the target is added at a concentration of 40 nM or lower (the actual concentration is kept at least 10-fold lower than the lowest protein dilution) and the reaction is allowed to equilibrate for at least 45 min. In addition the reaction mixture also contains 10 mM Tris (pH 7.5), 100 mM KCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 5 mM DTT, 10% glycerol, 0.02% BSA (poly (dIdC) or (dAdT) (Pharmacia) can also added at 10-100 μg/μl).

The equilibrated reactions are loaded onto a 10% polyacrylamide gel, which has been pre-run for 45 min in Tris/glycine buffer, then bound and unbound labeled target is resolved be electrophoresis at 150V (alternatively, 10-20% gradient Tris-HCl gels, containing a 4% polyacrylamide stacker, can be used). The dried gels are visualized by autoradiography or phosphoroimaging and the apparent K_d is determined by calculating the protein concentration that gives half-maximal binding.

filled in by high-fidelity thermostable polymerase, the combination of Taq and Pfu polymerases also suffices. In the second phase of construction, the zinc finger template is amplified by external primers designed to incorporate restriction sites at either end for cloning into a shuttle vector or directly into an expression vector.

An alternative method of cloning the newly designed DNA-binding proteins relies on annealing complementary oligonucleotides encoding the specific regions of the desired ZFP. This particular application requires that the oligonucleotides be phosphorylated prior to the final ligation step. This is usually performed before setting up the annealing reactions, but kinasing can also occur post-annealing. In brief, the "universal" oligonucleotides encoding the constant regions of the proteins (oligos 1, 2 and 3 of above) are annealed with their complementary oligonucleotides. Additionally, the "specific" oligonucleotides encoding the finger recognition helices are annealed with their respective complementary oligonucleotides. These complementary oligos are designed to fill in the region which was previously filled in by polymerase in the protocol described above. The complementary oligos to the common oligos 1 and finger 3 are engineered to leave overhanging sequences specific for the restriction sites used in cloning into the vector of choice. The second assembly protocol differs from the initial protocol in the following aspects: the "scaffold" encoding the newly designed ZFP is composed entirely of synthetic DNA thereby eliminating the polymerase fill-in step, additionally the fragment to be cloned into the vector does not require amplification. Lastly, the design of leaving sequence-specific overhangs eliminates the need for restriction enzyme digests of the inserting fragment.

The resulting fragment encoding the newly designed ZFP is ligated into an expression vector. Expression vectors that are commonly utilized include, but are not limited to, a modified pMAL-c2 bacterial expression vector (New England BioLabs, "NEB") or a eukaryotic expression vector, pcDNA (Promega).

Any suitable method of protein purification known to those of skill in the art can be used to purify ZFPs of the invention (*see* Ausubel, *supra*, Sambrook, *supra*). In addition, any suitable host can be used, e.g., bacterial cells, insect cells, yeast cells, mammalian cells, and the like.

In one embodiment, expression of the ZFP fused to a maltose binding protein (MBP-ZFP) in bacterial strain JM109 allows for straightforward purification through an amylose column (NEB). High expression levels of the zinc finger chimeric

specificity for a given target by an empirical process such as phage display. In some such methods, each component finger of a ZFP is designed or selected independently of other component fingers. For example, each finger can be obtained from a different preexisting ZFP or each finger can be subject to separate randomization and selection.

Once a ZFP has been selected, designed, or otherwise provided to a given target segment, the ZFP or the DNA encoding it are synthesized. Exemplary methods for synthesizing and expressing DNA encoding zinc finger proteins are described below. The ZFP or a polynucleotide encoding it can then be used for modulation of expression, or analysis of the target gene containing the target site to which the ZFP binds.

Expression and purification of ZFPs

ZFP polypeptides and nucleic acids can be made using routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)). In addition, essentially any nucleic acid can be custom ordered from any of a variety of commercial sources. Similarly, peptides and antibodies can be custom ordered from any of a variety of commercial sources.

Two alternative methods are typically used to create the coding sequences required to express newly designed DNA-binding peptides. One protocol is a PCR-based assembly procedure that utilizes six overlapping oligonucleotides (Figure 1). Three oligonucleotides (oligos 1, 3, and 5 in Figure 1) correspond to "universal" sequences that encode portions of the DNA-binding domain between the recognition helices. These oligonucleotides remain constant for all zinc finger constructs. The other three "specific" oligonucleotides (oligos 2, 4, and 6 in Figure 1) are designed to encode the recognition helices. These oligonucleotides contain substitutions primarily at positions -1, 2, 3 and 6 on the recognition helices making them specific for each of the different DNA-binding domains.

The PCR synthesis is carried out in two steps. First, a double stranded DNA template is created by combining the six oligonucleotides (three universal, three specific) in a four cycle PCR reaction with a low temperature annealing step, thereby annealing the oligonucleotides to form a DNA "scaffold." The gaps in the scaffold are

In a variation, the methods of the invention identify first and second target segments, each independently conforming to the above formula. The two target segments in such methods are constrained to be adjacent or proximate (i.e., within about 0-5 bases) of each other in the target gene. The strategy underlying selection of proximate target segments is to allow the design of a ZFP formed by linkage of two component ZFPs specific for the first and second target segments respectively. These principles can be extended to select target sites to be bound by ZFPs with any number of component fingers. For example, a suitable target site for a nine finger protein would have three component segments, each conforming to the above formula.

The target sites identified by the above methods can be subject to further evaluation by other criteria or can be used directly for design or selection (if needed) and production of a ZFP specific for such a site. A further criteria for evaluating potential target sites is their proximity to particular regions within a gene. If a ZFP is to be used to repress a cellular gene on its own (i.e., without linking the ZFP to a repressing moiety), then the optimal location appears to be at, or within 50 bp upstream or downstream of the site of transcription initiation, to interfere with the formation of the transcription complex (Kim & Pabo, *J. Biol. Chem.* 272:29795-296800 (1997)) or compete for an essential enhancer binding protein. If, however, a ZFP is fused to a functional domain such as the KRAB repressor domain or the VP16 activator domain, the location of the binding site is considerably more flexible and can be outside known regulatory regions. For example, a KRAB domain can repress transcription at a promoter up to at least 3 kbp from where KRAB is bound (Margolin *et al., PNAS* 91:4509-4513 (1994)). Thus, target sites can be selected that do not necessarily include or overlap segments of demonstrable biological significance with target genes, such as regulatory sequences. Other criteria for further evaluating target segments include the prior availability of ZFPs binding to such segments or related segments, and/or ease of designing new ZFPs to bind a given target segment.

After a target segment has been selected, a ZFP that binds to the segment can be provided by a variety of approaches. The simplest of approaches is to provide a precharacterized ZFP from an existing collection that is already known to bind to the target site. However, in many instances, such ZFPs do not exist. An alternative approach can also be used to design new ZFPs, which uses the information in a database of existing ZFPs and their respective binding affinities. A further approach is to design a ZFP based on substitution rules as discussed above. A still further alternative is to select a ZFP with

In the design of a ZFP with three fingers, a target site should be selected in which at least one finger of the protein, and optionally, two or all three fingers have the potential to bind a D-able site. Such can be achieved by selecting a target site from within a larger target gene having the formula 5'-NNx aNy bNzc-3', wherein

each of the sets (x, a), (y, b) and (z, c) is either (N, N) or (G, K);
at least one of (x, a), (y, b) and (z, c) is (G, K). and
N and K are IUPAC-IUB ambiguity codes

In other words, at least one of the three sets (x, a), (y, b) and (z, c) is the set (G, K), meaning that the first position of the set is G and the second position is G or T. Those of the three sets (if any) which are not (G, K) are (N, N), meaning that the first position of the set can be occupied by any nucleotide and the second position of the set can be occupied by any nucleotide. As an example, the set (x, a) can be (G, K) and the sets (y, b) and (z, c) can both be (N, N).

In the formula 5'-NNx aNy bNzc-3', the triplets of NNx aNy and bNzc represent the triplets of bases on the target strand bound by the three fingers in a ZFP. If only one of x, y and z is a G, and this G is followed by a K, the target site includes a single D-able subsite. For example, if only x is G, and a is K, the site reads 5'-NNG KNy bNzc-3' with the D-able subsite highlighted. If both x and y but not z are G, and a and b are K, then the target site has two overlapping D-able subsites as follows: 5'-NNG *KNG* *K*Nz c-3', with one such site being represented in bold and the other in italics. If all three of x, y and z are G and a, b, and c are K, then the target segment includes three D-able subsites, as follows 5'**NNG** ***KNG*** <u>***K***NG K</u>3', the D-able subsites being represented by bold, italics and underline.

These methods thus work by selecting a target gene, and systematically searching within the possible subsequences of the gene for target sites conforming to the formula 5'-NNx aNy bNzc-3', as described above. In some such methods, every possible subsequence of 10 contiguous bases on either strand of a potential target gene is evaluated to determine whether it conforms to the above formula, and, if so, how many D-able sites are present. Typically, such a comparison is performed by computer, and a list of target sites conforming to the formula are output. Optionally, such target sites can be output in different subsets according to how many D-able sites are present.

94:5525-5530 (1997); Griesman & Pabo, *Science* 275:657-661 (1997); Desjarlais & Berg, *PNAS* 91:11-99-11103 (1994)).

In a preferred embodiment, copending application USSN 09/229,007, filed January 12, 1999 provides methods that select a target gene, and identify a target site within the gene containing one to six (or more) D-able sites (see definition below). Using these methods, a ZFP can then be synthesized that binds to the preselected site. These methods of target site selection are premised, in part, on the recognition that the presence of one or more D-able sites in a target segment confers the potential for higher binding affinity in a ZFP selected or designed to bind to that site relative to ZFPs that bind to target segments lacking D-able sites (see below). Experimental evidence supporting this insight is provided in Examples 2-9 of copending application USSN 09/229,007, filed January 12, 1999.

A D-able site or subsite is a region of a target site that allows an appropriately designed single zinc finger to bind to four bases rather than three of the target site. Such a zinc finger binds to a triplet of bases on one strand of a double-stranded target segment (target strand) and a fourth base on the other strand (*see* Figure 2 of copending application USSN 09/229,007, filed January 12, 1999. Binding of a single zinc finger to a four base target segment imposes constraints both on the sequence of the target strand and on the amino acid sequence of the zinc finger. The target site within the target strand should include the "D-able" site motif 5' NNGK 3', in which N and K are conventional IUPAC-IUB ambiguity codes. A zinc finger for binding to such a site should include an arginine residue at position -1 and an aspartic acid, (or less preferably a glutamic acid) at position +2. The arginine residues at position -1 interacts with the G residue in the D-able site. The aspartic acid (or glutamic acid) residue at position +2 of the zinc finger interacts with the opposite strand base complementary to the K base in the D-able site. It is the interaction between aspartic acid (symbol D) and the opposite strand base (fourth base) that confers the name D-able site. As is apparent from the D-able site formula, there are two subtypes of D-able sites: 5' NNGG 3' and 5' NNGT 3'. For the former site, the aspartic acid or glutamic acid at position +2 of a zinc finger interacts with a C in the opposite strand to the D-able site. In the latter site, the aspartic acid or glutamic acid at position +2 of a zinc finger interacts with an A in the opposite strand to the D-able site. In general, NNGG is preferred over NNGT.

The following eight groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Glycine (G);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
7) Serine (S), Threonine (T); and
8) Cysteine (C), Methionine (M)

(*see, e.g.*, Creighton, *Proteins* (1984)).

Design of ZFPs

The ZFPs of the invention are engineered to recognize a selected target site in the endogenous gene of choice. Typically, a backbone from any suitable C_2H_2 ZFP, such as SP-1, SP-1C, or ZIF268, is used as the scaffold for the engineered ZFP (*see, e.g.*, Jacobs, *EMBO J.* 11:4507 (1992); Desjarlais & Berg, *PNAS* 90:2256-2260 (1993)). A number of methods can then be used to design and select a ZFP with high affinity for its target (e.g., preferably with a K_d of less than about 25 nM). As described above, a ZFP can be designed or selected to bind to any suitable target site in the target endogenous gene, with high affinity. Co-pending patent application USSN 09/229,007, filed January 12, 1999, comprehensively describes methods for design, construction, and expression of ZFPs for selected target sites.

Any suitable method known in the art can be used to design and construct nucleic acids encoding ZFPs, e.g., phage display, random mutagenesis, combinatorial libraries, computer/rational design, affinity selection, PCR, cloning from cDNA or genomic libraries, synthetic construction and the like. (*see, e.g.*, U.S. Pat. No. 5,786,538; Wu *et al.*, *PNAS* 92:344-348 (1995); Jamieson *et al.*, *Biochemistry* 33:5689-5695 (1994); Rebar & Pabo, *Science* 263:671-673 (1994); Choo & Klug, *PNAS* 91:11163-11167 (1994); Choo & Klug, *PNAS* 91: 11168-11172 (1994); Desjarlais & Berg, *PNAS* 90:2256-2260 (1993); Desjarlais & Berg, *PNAS* 89:7345-7349 (1992); Pomerantz *et al.*, *Science* 267:93-96 (1995); Pomerantz *et al.*, *PNAS* 92:9752-9756 (1995); and Liu *et al.*, *PNAS*

backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

cells, or mammalian cells such as CHO, HeLa, 293, COS-1, and the like, e.g., cultured cells (*in vitro*), explants and primary cultures (*in vitro* and *ex vivo*), and cells *in vivo*.

"Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms also apply to amino acid polymers in which one or more amino acid residues is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymer.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription. As used herein, a promoter typically includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of certain RNA polymerase II type promoters, a TATA element, enhancer, CCAAT box, SP-1 site, etc. As used herein, a promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. The promoters often have an element that is responsive to transactivation by a DNA-binding moiety such as a polypeptide, e.g., a nuclear receptor, Gal4, the lac repressor and the like.

A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under certain environmental or developmental conditions.

A "weak promoter" refers to a promoter having about the same activity as a wild type herpes simplex virus ("HSV") thymidine kinase ("tk") promoter or a mutated HSV tk promoter, as described in Eisenberg & McKnight, *Mol. Cell. Biol.* 5:1940-1947 (1985).

The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell, and optionally integration or replication of the expression vector in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment, of viral or non-viral origin. Typically, the expression vector includes an "expression cassette," which comprises a nucleic acid to be transcribed operably linked to a promoter. The term expression vector also encompasses naked DNA operably linked to a promoter.

By "host cell" is meant a cell that contains a ZFP or an expression vector or nucleic acid encoding a ZFP. The host cell typically supports the replication or expression of the expression vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, fungal, protozoal, higher plant, insect, or amphibian

Activators and repressors include co-activators and co-repressors (*see, e.g.*, Utley *et al.*, *Nature* 394:498-502 (1998)).

A "regulatory domain" refers to a protein or a protein domain that has transcriptional modulation activity when tethered to a DNA binding domain, i.e., a ZFP. Typically, a regulatory domain is covalently or non-covalently linked to a ZFP to effect transcription modulation. Alternatively, a ZFP can act alone, without a regulatory domain, to effect transcription modulation.

The term "heterologous" is a relative term, which when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, a nucleic acid that is recombinantly produced typically has two or more sequences from unrelated genes synthetically arranged to make a new functional nucleic acid, e.g., a promoter from one source and a coding region from another source. The two nucleic acids are thus heterologous to each other in this context. When added to a cell, the recombinant nucleic acids would also be heterologous to the endogenous genes of the cell. Thus, in a chromosome, a heterologous nucleic acid would include an non-native (non-naturally occurring) nucleic acid that has integrated into the chromosome, or a non-native (non-naturally occurring) extrachromosomal nucleic acid. In contrast, a naturally translocated piece of chromosome would not be considered heterologous in the context of this patent application, as it comprises an endogenous nucleic acid sequence that is native to the mutated cell.

Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (e.g., a "fusion protein," where the two subsequences are encoded by a single nucleic acid sequence). *See, e.g.*, Ausubel, *supra*, for an introduction to recombinant techniques.

The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (naturally occurring) form of the cell or express a second copy of a native gene that is otherwise normally or abnormally expressed, under expressed or not expressed at all.

"Inhibition of gene expression that prevents gene activation" refers to the ability of a zinc finger protein to block or prevent binding of an activator molecule.

Modulation can be assayed by determining any parameter that is indirectly or directly affected by the expression of the target gene. Such parameters include, e.g., changes in RNA or protein levels, changes in protein activity, changes in product levels, changes in downstream gene expression, changes in reporter gene transcription (luciferase, CAT, β-galactosidase, β-glucuronidase, GFP (*see, e.g.,* Mistili & Spector, *Nature Biotechnology* 15:961-964 (1997)); changes in signal transduction, phosphorylation and dephosphorylation, receptor-ligand interactions, second messenger concentrations (e.g., cGMP, cAMP, IP3, and Ca^{2+}), cell growth, and neovascularization. These assays can be *in vitro*, *in vivo*, and *ex vivo*. Such functional effects can be measured by any means known to those skilled in the art, e.g., measurement of RNA or protein levels, measurement of RNA stability, identification of downstream or reporter gene expression, e.g., via chemiluminescence, fluorescence, colorimetric reactions, antibody binding, inducible markers, ligand binding assays; changes in intracellular second messengers such as cGMP and inositol triphosphate (IP3); changes in intracellular calcium levels; cytokine release, and the like.

To determine the level of gene expression modulation by a ZFP, cells contacted with ZFPs are compared to control cells, e.g., without the zinc finger protein or with a non-specific ZFP, to examine the extent of inhibition or activation. Control samples are assigned a relative gene expression activity value of 100%. Modulation/inhibition of gene expression is achieved when the gene expression activity value relative to the control is about 80%, preferably 50% (i.e., 0.5x the activity of the control), more preferably 25%, more preferably 5-0%. Modulation/activation of gene expression is achieved when the gene expression activity value relative to the control is 110% , more preferably 150% (i.e., 1.5x the activity of the control), more preferably 200-500%, more preferably 1000-2000% or more.

A "transcriptional activator" and a "transcriptional repressor" refer to proteins or effector domains of proteins that have the ability to modulate transcription, as described above. Such proteins include, e.g., transcription factors and co-factors (e.g., KRAB, MAD, ERD, SID, nuclear factor kappa B subunit p65, early growth response factor 1, and nuclear hormone receptors, VP16, VP64), endonucleases, integrases, recombinases, methyltransferases, histone acetyltransferases, histone deacetylases etc.

The phrase "RNA polymerase pause site" is described in Uptain *et al.*, *Annu. Rev. Biochem.* 66:117-172 (1997).

"Humanized" refers to a non-human polypeptide sequence that has been modified to minimize immunoreactivity in humans, typically by altering the amino acid sequence to mimic existing human sequences, without substantially altering the function of the polypeptide sequence (*see, e.g.*, Jones *et al.*, *Nature* 321:522-525 (1986), and published UK patent application No. 8707252). Backbone sequences for the ZFPs are preferably be selected from existing human C_2H_2 ZFPs (e.g., SP-1). Functional domains are preferably selected from existing human genes, (e.g., the activation domain from the p65 subunit of NF-κB). Where possible, the recognition helix sequences will be selected from the thousands of existing ZFP DNA recognition domains provided by sequencing the human genome. As much as possible, domains will be combined as units from the same existing proteins. All of these steps will minimize the introduction of new junctional epitopes in the chimeric ZFPs and render the engineered ZFPs less immunogenic.

"Administering" an expression vector, nucleic acid, ZFP, or a delivery vehicle to a cell comprises transducing, transfecting, electroporating, translocating, fusing, phagocytosing, shooting or ballistic methods, etc., i.e., any means by which a protein or nucleic acid can be transported across a cell membrane and preferably into the nucleus of a cell.

A "delivery vehicle" refers to a compound, e.g., a liposome, toxin, or a membrane translocation polypeptide, which is used to administer a ZFP. Delivery vehicles can also be used to administer nucleic acids encoding ZFPs, e.g., a lipid:nucleic acid complex, an expression vector, a virus, and the like.

The terms "modulating expression" "inhibiting expression" and "activating expression" of a gene refer to the ability of a ZFP to activate or inhibit transcription of a gene. Activation includes prevention of transcriptional inhibition (i.e., prevention of repression of gene expression) and inhibition includes prevention of transcriptional activation (i.e., prevention of gene activation).

"Activation of gene expression that prevents repression of gene expression" refers to the ability of a zinc finger protein to block or prevent binding of a repressor molecule.

using an electrophoretic mobility shift assay ("EMSA"), as described in Example I and on page 14 of the present specification. Unless an adjustment is made for ZFP purity or activity, the K_d calculations made using the method of Example I may result in an underestimate of the true K_d of a given ZFP. Preferably, the K_d of a ZFP used to modulate transcription of an endogenous cellular gene is less than about 100 nM, more preferably less than about 75 nM, more preferably less than about 50 nM, most preferably less than about 25 nM.

An "endogenous cellular gene" refers to a gene that is native to a cell, which is in its normal genomic and chromatin context, and which is not heterologous to the cell. Such cellular genes include, e.g., animal genes, plant genes, bacterial genes, protozoal genes, fungal genes, mitrochondrial genes, and chloroplastic genes.

An "endogenous gene" refers to a microbial or viral gene that is part of a naturally occurring microbial or viral genome in a microbially or virally infected cell. The microbial or viral genome can be extrachromosomal or integrated into the host chromosome. This term also encompasses endogenous cellular genes, as described above.

A "native chromatin environment" refers to the naturally occurring, structural relationship of genomic DNA (e.g., bacterial, animal, fungal, plant, protozoal, mitochondrial, and chloroplastic) and DNA-binding proteins (e.g., histones and bacterial DNA binding protein II), which together form chromosomes. The endogenous cellular gene can be in a transcriptionally active or inactive state in the native chromatin environment.

A "developmentally silent gene" or an "inactive gene" refers to a gene whose expression is repressed or not activated, i.e., turned off, in certain cell types, during certain developmental stages of a cell type, or during certain time periods in a cell type. Examples of developmentally inactive genes include EPO, GATA, interleukin family proteins, GM-CSF, MyoD, eutrophin, and fetal hemoglobins gamma and delta.

The phrase "adjacent to a transcription initiation site" refers to a target site that is within about 50 bases either upstream or downstream of a transcription initiation site. "Upstream" of a transcription initiation site refers to a target site that is more than about 50 bases 5' of the transcription initiation site (i.e., in the non-transcribed region of the gene).

The term "zinc finger protein" or "ZFP" refers to a protein having DNA binding domains that are stabilized by zinc. The individual DNA binding domains are typically referred to as "fingers" A zinc finger protein has at least one finger, typically two fingers, three fingers, four fingers, five fingers, or six fingers or more. Each finger binds from two to four base pairs of DNA, typically three or four base pairs of DNA. A zinc finger protein binds to a nucleic acid sequence called a target site or target segment. Each finger typically comprises an approximately 30 amino acid, zinc-coordinating, DNA-binding subdomain. An exemplary motif characterizing one class of these proteins (Cys_2His_2 class) is -Cys-$(X)_{2-4}$-Cys-$(X)_{12}$-His-$(X)_{3-5}$-His (where X is any amino acid). Studies have demonstrated that a single zinc finger of this class consists of an alpha helix containing the two invariant histidine residues co-ordinated with zinc along with the two cysteine residues of a single beta turn (*see, e.g.*, Berg & Shi, *Science* 271:1081-1085 (1996)).

A "target site" is the nucleic acid sequence recognized by a zinc finger protein. A single target site typically has about four to about ten or more base pairs. Typically, a two-fingered zinc finger protein recognizes a four to seven base pair target site, a three-fingered zinc finger protein recognizes a six to ten base pair target site, a six fingered zinc finger protein recognizes two adjacent nine to ten base pair target sites, and so on for proteins with more than six fingers. The target site is in any position that allows regulation of gene expression, e.g., adjacent to, up- or downstream of the transcription initiation site; proximal to an enhancer or other transcriptional regulation element such as a repressor (e.g., SP-1 binding sites, hypoxia response elements, nuclear receptor recognition elements, p53 binding sites, etc.), RNA polymerase pause sites; and intron/exon boundaries. The term "adjacent target sites" refers to non-overlapping target sites that are separated by zero to about 5 base pairs.

"K_d" refers to the dissociation constant for the compound, i.e., the concentration of a compound (e.g., a zinc finger protein) that gives half maximal binding of the compound to its target (i.e., half of the compound molecules are bound to the target) under given conditions (i.e., when [target] << K_d), as measured using a given assay system (*see, e.g.*, U.S. Patent No. 5,789,538). The assay system used to measure the K_d should be chosen so that it gives the most accurate measure of the actual K_d of the ZFP. Any assay system can be used, as long is it gives an accurate measurement of the actual K_d of the ZFP. In one embodiment, the K_d for the ZFPs of the invention is measured

target site is expected to provide specificity in the human genome, as a target site of that size should occur only once in every $3x10^{10}$ base pairs, and the size of the human genome is $3.5x10^9$ base pairs (*see, e.g.*, Liu *et al.*, *PNAS* 94:5525-5530 (1997)). In another embodiment, the ZFPs are non-covalently associated, through a leucine zipper, a STAT protein N-terminal domain, or the FK506 binding protein (*see, e.g.*, O'Shea, *Science* 254: 539 (1991), Barahmand-Pour *et al.*, *Curr. Top. Microbiol. Immunol.* 211:121-128 (1996); Klemm *et al.*, *Annu. Rev. Immunol.* 16:569-592 (1998); Ho *et al.*, *Nature* 382:822-826 (1996)).

In another embodiment, the ZFP is linked to at least one or more regulatory domains, described below. Preferred regulatory domains include transcription factor repressor or activator domains such as KRAB and VP16, co-repressor and co-activator domains, DNA methyl transferases, histone acetyltransferases, histone deacetylases, and endonucleases such as Fok1. For repression of gene expression, typically the expression of the gene is reduced by about 20% (i.e., 80% of non-ZFP modulated expression), more preferably by about 50% (i.e., 50% of non-ZFP modulated expression), more preferably by about 75-100% (i.e., 25% to 0% of non-ZFP modulated expression). For activation of gene expression, typically expression is activated by about 1.5 fold (i.e., 150% of non-ZFP modulated expression), preferably 2 fold (i.e., 200% of non-ZFP modulated expression), more preferably 5-10 fold (i.e., 500-1000% of non-ZFP modulated expression), up to at least 100 fold or more.

The expression of engineered ZFP activators and repressors can be also controlled by systems typified by the tet-regulated systems and the RU-486 system (*see, e.g.*, Gossen & Bujard, *PNAS* 89:5547 (1992); Oligino *et al.*, *Gene Ther.* 5:491-496 (1998); Wang *et al.*, *Gene Ther.* 4:432-441 (1997); Neering *et al.*, *Blood* 88:1147-1155 (1996); and Rendahl *et al.*, *Nat. Biotechnol.* 16:757-761 (1998)). These impart small molecule control on the expression of the ZFP activators and repressors and thus impart small molecule control on the target gene(s) of interest. This beneficial feature could be used in cell culture models, in gene therapy, and in transgenic animals and plants.

Definitions

As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

desaturase, delta-9 desaturase, delta-15 desaturase, acetyl-CoA carboxylase, acyl-ACP-thioesterase, ADP-glucose pyrophosphorylase, starch synthase, cellulose synthase, sucrose synthase, senescence-associated genes, heavy metal chelators, fatty acid hydroperoxide lyase, viral genes, protozoal genes, fungal genes, and bacterial genes. In general, suitable genes to be regulated include cytokines, lymphokines, growth factors, mitogenic factors, chemotactic factors, onco-active factors, receptors, potassium channels, G-proteins, signal transduction molecules, and other disease-related genes. Preferred developmentally inactive genes include EPO, GATA, interleukin family proteins, GM-CSF, MyoD, eutrophin, and fetal hemoglobins gamma and delta.

A general theme in transcription factor function is that simple binding and sufficient proximity to the promoter are all that is generally needed. Exact positioning relative to the promoter, orientation, and within limits, distance, do not matter greatly for expression modulation by a ZFP. This feature allows considerable flexibility in choosing sites for constructing artificial transcription factors. The target site recognized by the ZFP therefore can be any suitable site in the target gene that will allow activation or repression of gene expression by a ZFP, optionally linked to a regulatory domain. Preferred target sites include regions adjacent to, downstream, or upstream of the transcription start site. In addition, target sites can also be located in enhancer regions, repressor sites, RNA polymerase pause sites, and specific regulatory sites (e.g., SP-1 sites, hypoxia response elements, nuclear receptor recognition elements, p53 binding sites), sites in the cDNA encoding region or in an expressed sequence tag (EST) coding region. As described below, typically each finger recognizes 2-4 base pairs, with a two finger ZFP binding to a 4 to 7 bp target site, a three finger ZFP binding to a 6 to 10 base pair site, and a six finger ZFP binding to two adjacent target sites, each target site having from 6-10 base pairs.

As described herein, two ZFPs can be administered to a cell, recognizing either the same target endogenous cellular gene, or different target endogenous cellular gene. The first ZFP optionally is associated with the second ZFP, either covalently or non-covalently. Recognition of adjacent target sites by either associated or individual ZFPs can be used to produce cooperative binding of the ZFPs, resulting in an affinity that is greater than the affinity of the ZFPs when individually bound to their target site.

In one embodiment, two ZFPs are produced as a fusion protein linked by an amino acid linker, and the resulting six finger ZFP recognizes an approximately 18 base pair target site (*see, e.g.*, Liu *et al., PNAS* 94:5525-5530 (1997)). An 18 base pair

chromatin environment. The present invention thus provides zinc finger DNA binding proteins that have been engineered to specifically recognize, with high efficacy, endogenous cellular genes. The experiments described herein demonstrate that a 3 finger ZFP with a target site affinity of less than about 10 nM (VEGF1) can be used to effectively activate or repress activity of an endogenous gene. Furthermore, a 6 finger ZFP (VEGF3a/1) was also shown to effectively repress activity of an endogenous gene. Finally, three finger ZFP can be used to activate endogenous EPO, a developmentally inactive gene. Preferably, the ZFPs of the invention exhibit high affinity for their target sites, with K_ds of less than about 100 nM, preferably less than about 50 nM, most preferably less than about 25 nM or lower.

As a result, the ZFPs of the invention can be used to regulate endogenous gene expression, both through activation and repression of endogenous gene transcription. The ZFPs can also be linked to regulatory domains, creating chimeric transcription factors to activate or repress transcription. In one preferred embodiment, the methods of regulation use ZFPs with a K_d of less than about 25 nM to activate or repress gene transcription. The ZFPs of the invention therefore can be used to repress transcription of an endogenous cellular gene by 20% or more, and can be used to activate transcription of an endogenous cellular gene by about 1.5 fold or more.

Such methods of regulating gene expression allow for novel human and mammalian therapeutic applications, e.g., treatment of genetic diseases, cancer, fungal, protozoal, bacterial, and viral infection, ischemia, vascular disease, arthritis, immunological disorders, etc., as well as providing means for functional genomics assays, and means for developing plants with altered phenotypes, including disease resistance, fruit ripening, sugar and oil composition, yield, and color.

As described herein, ZFPs can be designed to recognize any suitable target site, for regulation of expression of any endogenous gene of choice. Examples of endogenous genes suitable for regulation include VEGF, CCR5, ERα, Her2/Neu, Tat, Rev, HBV C, S, X, and P, LDL-R, PEPCK, CYP7, Fibrinogen, ApoB, Apo E, Apo(a), renin, NF-κB, I-κB, TNF-α, FAS ligand, amyloid precursor protein, atrial naturetic factor, ob-leptin, ucp-1, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-12, G-CSF, GM-CSF, Epo, PDGF, PAF, p53, Rb, fetal hemoglobin, dystrophin, eutrophin, GDNF, NGF, IGF-1, VEGF receptors flt and flk, topoisomerase, telomerase, bcl-2, cyclins, angiostatin, IGF, ICAM-1, STATS, c-myc, c-myb, TH, PTI-1, polygalacturonase, EPSP synthase, FAD2-1, delta-12

Figure 7. Co-transfection data showing activation of luciferase reporter activity via VEGF-VP16 protein expression. Error bars show the standard deviation of triplicate transfections. pGL3-P (reporter with no VEGF target); pcDNA (empty effector vector control); pVFR3-4x (VEGF reporter plasmid); VEGF1 (VEGF1-VP16); VEGF3a (VEGF3a-VP16); VEGF3a/1 (VEGF3a/1-VP16).

Figure 8. VEGF ELISA data showing repression of endogenous VEGF gene expression due to transfection of a VEGF ZFP-KRAB effector plasmid. DFX treated (control nontransfected Dfx treated cells; No ZFP (pcDNA-control), VEGF 1 (VEGF1-KRAB), VEGF 3a/1 (VEGF3a/1-KRAB), CCR5 (CCR5-KRAB); Mock uninduced (mock transfected cells untreated with DFX). Error bars show the standard deviation of duplicate transfections.

Figure 9. VEGF ELISA data showing activation of endogenous VEGF gene expression due to transfection of a VEGF ZFP-VP16 effector plasmid. Mock (mock transfected cells); No ZFP (NVF-control), VEGF 1 (VEGF1-VP16), VEGF 3a/1 (VEGF3a/1-VP16). Error bars show the standard deviation of duplicate transfections.

Figure 10. RNase protection assay showing changes in VEGF specific mRNA by VEGF-specific ZFPs. Panel A: Activation of VEGF mRNA, NVF-Control (no ZFP), VEGF1-NVF (VEGF1-VP16), CCR5-5-NVF (CCR5-VP16), CCR5-3-NVF (CCR5-VP16). Panel B: Repression of VEGF mRNA. NKF-Control (no ZFP), VEGF1-NKF (VEGF1-KRAB), VEGF3a/1-NKF (VEGF3a/1-KRAB), CCR5-3-NKF (CCR5-KRAB). The size of the 148 nucleotide VEGF specific band is indicated by an arrow. The VEGF specific probe was synthesized from a human angiogenesis multi-probe template set (Pharmingen).As a control, signals from the housekeeping genes L32 and GAPDH are shown (arrows).

Figure 11a-b: Figure 11a shows activation of endogenous EPO gene expression by measuring EPO production in Hep3B and 293 cells, as measuring using ELISA. Figure 11b shows activation of endogenous EPO gene expression in Hep3B cells and 293 cells by measuring mRNA expression.

DETAILED DESCRIPTION OF THE INVENTION

Introduction

The present application demonstrates for the first time that ZFPs can be used to regulate expression of an endogenous cellular gene that is present in its native

target site is adjacent to an RNA polymerase pause site downstream of a transcription initiation site of the endogenous cellular gene.

In another embodiment, the zinc finger protein comprises an SP-1 backbone. In one embodiment, the zinc finger protein comprises a regulatory domain and is humanized.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: PCR amplification scheme for production of ZFP-encoding synthetic genes.

Figure 2. Expression and purification of typical ZFPs. Fig. 2A: Unfused ZFP before induction (lane 1), after induction (lane 2), and after purification (lane 3). Fig. 2B: MBP-VEGF expression before induction (lane 1), after induction (lane 2), and after French Press lysis (lane 3). Fig. 2C: Purification of MBP-VEGF by amylose affinity column showing flow-through (FT), and initial fractions (1-4). Fraction 2 was used for electrophoretic mobility shift assays ("EMSA"). M, molecular weight markers.

Figure 3. Typical EMSA experiment with MBP fused ZFP. MBP-VEGF1 protein was bound to labeled duplex DNA as described in the text. A three-fold protein dilution series was carried out; each point represents the percent shifted at that particular protein concentration plotted on a semi-log graph. Quantitation was by phosphorimager. In this case, the protein concentration yielding 50% of maximum shift (the apparent K_d) was 2 nM.

Figure 4. Off-rate experiment comparing VEGF1 to VEGF3a/1. Protein-DNA complexes were pre-formed and incubated with a 1000-fold excess of unlabeled oligonucleotide. Samples were electrophoresed at various times and the amount of shifted product was measured by phosphorimager. Curve fitting was used to calculate the indicated complex half-lives.

Figure 5. Typical expression vector used for transient ZFP expression in mammalian cells.

Figure 6. Co-transfection data showing repression of luciferase reporter activity via VEGF-KRAB protein expression. Error bars show the standard deviation of triplicate transfections. pGL3-C (reporter vector control); pVFR1-4x (VEGF reporter plasmid); VEGF1 (VEGF1-KRAB); VEGF3a (VEGF3a-KRAB); VEGF3a/1 (VEGF3a/1-KRAB).

In one embodiment, the endogenous cellular gene is a selected from the group consisting of VEGF, ERα, IGF-I, c-myc, c-myb, ICAM, Her2/Neu, FAD2-1, EPO, GM-CSF, GDNF, and LDL-R. In another embodiment, the endogenous cellular gene is a developmentally silent or otherwise inactive gene, e.g., EPO, GATA, interleukin family proteins, GM-CSF, MyoD, eutrophin, and fetal hemoglobins gamma and delta. In another embodiment, the regulatory domain is selected from the group consisting of a transcriptional repressor, a transcriptional activator, an endonuclease, a methyl transferase, a histone acetyltransferase, and a histone deacetylase.

In one embodiment, the cell is selected from the group consisting of animal cell, a plant cell, a bacterial cell, a protozoal cell, or a fungal cell. In another embodiment, the cell is a mammalian cell. In another embodiment, the cell is a human cell.

In one embodiment, the method further comprises the step of first administering to the cell a delivery vehicle comprising the zinc finger protein, wherein the delivery vehicle comprises a liposome or a membrane translocation polypeptide.

In one embodiment, the zinc finger protein is encoded by a zinc finger protein nucleic acid operably linked to a promoter, and the method further comprises the step of first administering the nucleic acid to the cell in a lipid:nucleic acid complex or as naked nucleic acid. In another embodiment, the zinc finger protein is encoded by an expression vector comprising a zinc finger protein nucleic acid operably linked to a promoter, and the method further comprises the step of first administering the expression vector to the cell. In another embodiment, the expression vector is a viral expression vector. In another embodiment, the expression vector is a retroviral expression vector, an adenoviral expression vector, a DNA plasmid expression vector, or an AAV expression vector.

In one the zinc finger protein is encoded by a nucleic acid operably linked to an inducible promoter. In another embodiment, the zinc finger protein is encoded by a nucleic acid operably linked to a weak promoter.

In one embodiment, the cell comprises less than about 1.5×10^6 copies of the zinc finger protein.

In one embodiment, the target site is upstream of a transcription initiation site of the endogenous cellular gene. In another embodiment, the target site is adjacent to a transcription initiation site of the endogenous cellular gene. In another embodiment, the

the K_d of the zinc finger protein is less than about 25 nM; thereby activating expression of the endogenous cellular gene to at least about 150%.

In another aspect, the present invention provides a method of activating expression of an endogenous cellular gene, the method comprising the step of: contacting a target site in the endogenous cellular gene with a fusion zinc finger protein comprising six fingers and a regulatory domain, wherein the K_d of the zinc finger protein is less than about 25 nM; thereby activating expression of the endogenous cellular gene to at least about 150%.

In one embodiment, expression of the endogenous cellular gene is activated to at least about 200-500%. In another embodiment, activation of gene expression prevents repression of gene expression.

In another aspect, the present invention provides a method of modulating expression of an endogenous cellular gene in a cell, the method comprising the step of: contacting a first target site in the endogenous cellular gene with a first zinc finger protein; thereby modulating expression of the endogenous cellular gene.

In one embodiment, the zinc finger protein has two, three, four, five, or six fingers.

In another aspect, the present invention provides a method of modulating expression of an endogenous cellular gene in a cell, the method comprising the step of: contacting a target site in the endogenous cellular gene with a fusion zinc finger protein comprising six fingers and a regulatory domain; thereby modulating expression of the endogenous cellular gene.

In one embodiment, the step of contacting further comprises contacting a second target site in the endogenous cellular gene with a second zinc finger protein. In another embodiment, the first and second target sites are adjacent. In another embodiment, the first and second zinc finger proteins are covalently linked. In another embodiment, the first zinc finger protein is a fusion protein comprising a regulatory domain. In another embodiment, the first zinc finger protein is a fusion protein comprising at least two regulatory domains. In another embodiment, the first and second zinc finger proteins are fusion proteins, each comprising a regulatory domain. In another embodiment, the first and second zinc finger protein are fusion proteins, each comprising at least two regulatory domains.

is randomly integrated into the genome and is not found in a native chromatin environment as compared to an endogenous gene. In contrast, specific regulation of an endogenous cellular gene in its native chromatin environment using a ZFP has not yet been demonstrated in the art.

SUMMARY OF THE INVENTION

The present invention thus provides for the first time methods of regulating endogenous cellular gene expression, where the endogenous genes are in their native chromatin environment, in contrast to genes that have been transiently expressed in a cell, or those that have been exogenously integrated into the genome. In addition, the present invention provides for the first time activation of a developmentally silent, endogenous gene. In one preferred embodiment, the methods of regulation use ZFPs with a K_d of less than about 25 nM to activate or repress gene transcription. The ZFPs of the invention therefore can be used to repress transcription of an endogenous cellular gene by 20% or more, and can be used to activate transcription of an endogenous cellular gene by about 1.5 fold or more.

In one aspect, the present invention provides a method of inhibiting expression of an endogenous cellular gene in a cell, the method comprising the step of: contacting a first target site in the endogenous cellular gene with a first zinc finger protein, wherein the K_d of the zinc finger protein is less than about 25 nM; thereby inhibiting expression of the endogenous cellular gene by at least about 20%.

In another aspect, the present invention provides a method of inhibiting expression of an endogenous cellular gene in a cell, the method comprising the step of: contacting a target site in the endogenous cellular gene with a fusion zinc finger protein comprising six fingers and a regulatory domain, wherein the K_d of the zinc finger protein is less than about 25 nM; thereby inhibiting expression of the endogenous cellular gene by at least about 20%.

In one embodiment, expression of the endogenous cellular gene is inhibited by at least about 75%-100%. In another embodiment, the inhibition of gene expression prevents gene activation.

In another aspect, the present invention provides a method of activating expression of an endogenous cellular gene, the method comprising the step of: contacting a first target site in the endogenous cellular gene with a first zinc finger protein, wherein

its two components. The authors then constructed a reporter vector containing a luciferase gene operably linked to a promoter and a hybrid site for the chimeric DNA binding protein in proximity to the promoter. The authors reported that their chimeric DNA binding protein could activate or repress expression of the luciferase gene.

Liu *et al.*, *PNAS* 94:5525-5530 (1997) report forming a composite ZFP by using a peptide spacer to link two component ZFPs, each having three fingers. The composite protein was then further linked to transcriptional activation or repression domains. It was reported that the resulting chimeric protein bound to a target site formed from the target segments bound by the two component ZFPs. It was further reported that the chimeric ZFP could activate or repress transcription of a reporter gene when its target site was inserted into a reporter plasmid in proximity to a promoter operably linked to the reporter.

Beerli *et al.*, *PNAS* 95:14628-14633 (1998) report construction of a chimeric six finger ZFP fused to either a KRAB, ERD, or SID transcriptional repressor domain, or the VP16 or VP64 transcriptional activation domain. This chimeric ZFP was designed to recognize an 18 bp target site in the 5' untranslated region of the human erbB-2 gene. Using this construct, the authors of this study report both activation and repression of a transiently expressed reporter luciferase construct linked to the erbB-2 promoter.

In addition, a recombinant ZFP was reported to repress expression of an integrated plasmid construct encoding a *bcr-abl* oncogene (Choo *et al.*, *Nature* 372:642-645 (1994)). The target segment to which the ZFPs bound was a nine base sequence GCA GAA GCC chosen to overlap the junction created by a specific oncogenic translocation fusing the genes encoding bcr and abl. The intention was that a ZFP specific to this target site would bind to the oncogene without binding to abl or bcr component genes. The authors used phage display to select a variant ZFP that bound to this target segment. The variant ZFP thus isolated was then reported to repress expression of a stably transfected *bcr-abl* construct in a cell line.

To date, these methods have focused on regulation of either transiently expressed genes, or on regulation of exogenous genes that have been integrated into the genome. The transiently expressed genes described by Pomerantz *et al.*, Liu *et al.*, and Beerli *et al.* are episomal and are not packaged into chromatin in the same manner as chromosomal genes. Moreover, even the stably expressed gene described by Choo *et al.*

shows that each finger can be superimposed on the next by a periodic rotation and translation of the finger along the main DNA axis. The structure suggests that each finger interacts independently with DNA over 3 base-pair intervals, with side-chains at positions -1, 2 , 3 and 6 on each recognition helix making contacts with respective DNA triplet sub-site. The amino terminus of Zif268 is situated at the 3' end of its DNA recognition subsite. Some zinc fingers can bind to a fourth base in a target segment. The fourth base is on the opposite strand from the other three bases recognized by zinc finger and complementary to the base immediately 3' of the three base subsite.

The structure of the Zif268-DNA complex also suggested that the DNA sequence specificity of a ZFP might be altered by making amino acid substitutions at the four helix positions (-1, 2, 3 and 6) on a zinc finger recognition helix. Phage display experiments using zinc finger combinatorial libraries to test this observation were published in a series of papers in 1994 (Rebar *et al.*, *Science* 263:671-673 (1994); Jamieson *et al.*, *Biochemistry* 33:5689-5695 (1994); Choo *et al.*, *PNAS* 91:11163-11167 (1994)). Combinatorial libraries were constructed with randomized side-chains in either the first or middle finger of Zif268 and then isolated with an altered Zif268 binding site in which the appropriate DNA sub-site was replaced by an altered DNA triplet. Correlation between the nature of introduced mutations and the resulting alteration in binding specificity gave rise to a partial set of substitution rules for rational design of ZFPs with altered binding specificity.

Greisman & Pabo, *Science* 275:657-661 (1997) discuss an elaboration of a phage display method in which each finger of a zinc finger protein is successively subjected to randomization and selection. This paper reported selection of ZFPs for a nuclear hormone response element, a p53 target site and a TATA box sequence.

Recombinant ZFPs have been reported to have the ability to regulate gene expression of transiently expressed reporter genes in cultured cells (*see, e.g.*, Pomerantz *et al.*, *Science* 267:93-96 (1995); Liu *et al.*, *PNAS* 94:5525-5530 1997); and Beerli *et al.*, *PNAS* 95:14628-14633 (1998)).

For example, Pomerantz *et al.*, *Science* 267:93-96 (1995) report an attempt to design a novel DNA binding protein by fusing two fingers from Zif268 with a homeodomain from Oct-1. The hybrid protein was then fused with either a transcriptional activator or repressor domain for expression as a chimeric protein. The chimeric protein was reported to bind a target site representing a hybrid of the subsites of

Gene expression is normally controlled through alterations in the function of sequence specific DNA binding proteins called transcription factors. These bind in the general proximity (although occasionally at great distances) of the point of transcription initiation of a gene. They act to influence the efficiency of formation or function of a transcription initiation complex at the promoter. Transcription factors can act in a positive fashion (transactivation) or in a negative fashion (transrepression).

Transcription factor function can be constitutive (always "on") or conditional. Conditional function can be imparted on a transcription factor by a variety of means, but the majority of these regulatory mechanisms depend of the sequestering of the factor in the cytoplasm and the inducible release and subsequent nuclear translocation, DNA binding and transactivation (or repression). Examples of transcription factors that function this way include progesterone receptors, sterol response element binding proteins (SREBPs) and NF-kappa B. There are examples of transcription factors that respond to phosphorylation or small molecule ligands by altering their ability to bind their cognate DNA recognition sequence (Hou *et al.*, *Science* 256:1701 (1994); Gossen & Bujard, *PNAS* 89:5547 (1992); Oligino *et al.*, *Gene Ther.* 5:491-496 (1998); Wang *et al.*, *Gene Ther.* 4:432-441 (1997); Neering *et al.*, *Blood* 88:1147-1155 (1996); and Rendahl *et al.*, *Nat. Biotechnol.* 16:757-761 (1998)). This mechanism is common in prokaryotes but somewhat less common in eukaryotes.

Zinc finger proteins ("ZFPs") are proteins that can bind to DNA in a sequence-specific manner. Zinc fingers were first identified in the transcription factor TFIIIA from the oocytes of the African clawed toad, *Xenopus laevis*. ZFPs are widespread in eukaryotic cells. An exemplary motif characterizing one class of these proteins (C_2H_2 class) is -Cys-$(X)_{2-4}$-Cys-$(X)_{12}$-His-$(X)_{3-5}$-His (where X is any amino acid). A single finger domain is about 30 amino acids in length and several structural studies have demonstrated that it contains an alpha helix containing the two invariant histidine residues co-ordinated through zinc with the two cysteines of a single beta turn. To date, over 10,000 zinc finger sequences have been identified in several thousand known or putative transcription factors. ZFPs are involved not only in DNA-recognition, but also in RNA binding and protein-protein binding. Current estimates are that this class of molecules will constitute about 2% of all human genes.

The X-ray crystal structure of Zif268, a three-finger domain from a murine transcription factor, has been solved in complex with its cognate DNA-sequence and

25:3559(1997); Wolffe *et al., PNAS 96:5894(1999)*). Examples of the therapeutic benefit for expression of such genes include activation of developmentally silent fetal hemoglobin genes to treat sickle cell disease and the activation of eutrophin to treat muscular dystrophy. In addition, pathogenic organisms such as viruses, bacteria, fungi, and protozoa could be controlled by altering gene expression. There is thus a clear unmet need for therapeutic approaches that act through sequence-specific regulation of disease-related genes.

In addition to the direct therapeutic utility provided by the ability to manipulate gene expression, this ability can be used experimentally to determine the function of a gene of interest. One common existing method for experimentally determining the function of a newly discovered gene is to clone its cDNA into an expression vector driven by a strong promoter and measure the physiological consequence of its over-expression in a transfected cell. This method is labor intensive and does not address the physiological consequences of down-regulation of a target gene. Simple methods allowing the selective over and under-expression of uncharacterized genes would be of great utility to the scientific community. Methods that permit the regulation of genes in cell model systems, transgenic animals and transgenic plants would find widespread use in academic laboratories, pharmaceutical companies, genomics companies and in the biotechnology industry.

An additional use of tools permitting the manipulation of gene expression is in the production of commercially useful biological products. Cell lines, transgenic animals and transgenic plants could be engineered to over-express a useful protein product. The production of erythropoietin by such an engineered cell line serves as an example. Likewise, production from metabolic pathways might be altered or improved by the selective up or down-regulation of a gene encoding a crucial enzyme. An example of this is the production of plants with altered levels of fatty acid saturation.

Methods currently exist in the art, which allow one to alter the expression of a given gene, e.g., using ribozymes, antisense technology, small molecule regulators, over-expression of cDNA clones, and gene-knockouts. These methods have to date proven to be generally insufficient for many applications and typically have not demonstrated either high target efficacy or high specificity *in vivo*. For useful experimental results and therapeutic treatments, these characteristics are desired.

REGULATION OF ENDOGENOUS GENE EXPRESSION IN CELLS USING ZINC FINGER PROTEINS

CROSS-REFERENCES TO RELATED APPLICATIONS

This application is a continuation-in-part of and claims the benefit of USSN 09/229,037, filed January 12, 1999. This application is also related to USSN 09/229,007, filed January 12, 1999, herein both incorporated by reference in their entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This invention was made with government support under Grant No. 1 R43 DK52251-01, awarded by the National Institutes of Health. The government has certain rights in this invention.

FIELD OF THE INVENTION

The present invention provides methods for regulating gene expression of endogenous genes using recombinant zinc finger proteins.

BACKGROUND OF THE INVENTION

Many pathophysiological processes are the result of aberrant gene expression. Examples include the inappropriate activation of proinflammatory cytokines in rheumatoid arthritis, under-expression of the hepatic LDL receptor in hypercholesteremia, over-expression of proangiogenic factors, and under-expression of antiangiogenic factors in solid tumor growth. If therapeutic methods for control of gene expression existed, many of these pathologies could be more optimally treated. In another example, developmentally silent or otherwise inactive genes could be activated in order to treat a particular disease state. Inactive genes are repressed via several mechanisms, including chromatin structure, specific *cis*-acting repressors, and DNA methylation (Travers, *Cell* 96:311(1996); Beato & Eisfeld, *Nucleic Acids Res.*

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68 — A1

(11) International Publication Number: WO 00/42219

(43) International Publication Date: 20 July 2000 (20.07.00)

(21) International Application Number: PCT/US00/00388

(22) International Filing Date: 6 January 2000 (06.01.00)

(30) Priority Data:
09/229,007 12 January 1999 (12.01.99) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 09/229,007 (CON)
Filed on 12 January 1999 (12.01.99)

(71) Applicant *(for all designated States except US)*: SANGAMO BIOSCIENCES, INC. [US/US]; Point Richmont Tech Center, Suite A100, 501 Canal Boulevard, Richmond, CA 94804 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: EISENBERG, Stephen, P. [US/US]; 5664 Pennsylvania Place, Boulder, CO 80303 (US). CASE, Casey, C. [US/US]; 605 Barroilhet Avenue, San Mateo, CA 94402 (US). COX, George, N., III [US/US]; 678 West Willow Street, Louisville, CO 80027 (US). JAMIESON, Andrew [GB/US]; 2528 Sutter Street, San Francisco, CA 94115 (US). REBAR, Edward, J. [US/US]; 815 Regal Road, Berkeley, CA 94708 (US).

(74) Agents: LIEBESCHUETZ, Joe et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, 8th floor, San Francisco, CA 94111-3834 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: SELECTION OF SITES FOR TARGETING BY ZINC FINGER PROTEINS AND METHODS OF DESIGNING ZINC FINGER PROTEINS TO BIND TO PRESELECTED SITES

(57) Abstract

The invention provides criteria and methods for selecting optimum subsequence(s) from a target for targeting by a zinc finger protein. Some of the methods of targetgene site selection seek to identify one or more target segments having a DNA motif containing one or more so-called D-able subsites having the sequence 5'NNGK3'. Other methods of the invention are directed to selection of target segments within target genes using a correspondence regime between different triplets of three bases and the three possible positions of a triplet within a nine-base site. In another aspect, the invention provides methods of designing zinc finger proteins that bind to a preselected target site. These methods can be used following the preselection of target sites according to the procedures and criteria described above. The methods of design use a database containing information about previously characterized zinc finger proteins.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SELECTION OF SITES FOR TARGETING BY ZINC FINGER PROTEINS AND METHODS OF DESIGNING ZINC FINGER PROTEINS TO BINDS TO PRESELECTED SITES

TECHNICAL FIELD

The invention resides in the technical fields of bioinformatics, and protein engineering.

BACKGROUND

Zinc finger proteins (ZFPs)are proteins that can bind to DNA in a sequence-specific manner. Zinc fingers were first identified in the transcription factor TFIIIA from the oocytes of the African clawed toad, *Xenopus laevis*. An exemplary motif characterizing one class of these protein (C_2H_2 class) is -Cys-$(X)_{2-4}$-Cys-$(X)_{12}$-His-$(X)_{3-5}$-His (where X is any amino acid). A single finger domain is about 30 amino acids in length, and several structural studies have demonstrated that it contains an alpha helix containing the two invariant histidine residues and two invariant cysteine residues in a beta turn co-ordinated through zinc. To date, over 10,000 zinc finger sequences have been identified in several thousand known or putative transcription factors. Zinc finger domains are involved not only in DNA-recognition, but also in RNA binding and in protein-protein binding. Current estimates are that this class of molecules will constitute about 2% of all human genes.

The x-ray crystal structure of Zif268, a three-finger domain from a murine transcription factor, has been solved in complex with a cognate DNA-sequence and shows that each finger can be superimposed on the next by a periodic rotation. The structure suggests that each finger interacts independently with DNA over 3 base-pair intervals, with side-chains at positions -1, 2 , 3 and 6 on each recognition helix making contacts with their respective DNA triplet subsites. The amino terminus of Zif268 is situated at the 3' end of the DNA strand with which it makes most contacts. Recent results have indicated that some zinc fingers can bind to a fourth base in a target segment.

If the strand with which a zinc finger protein makes most contacts is designated the target strand, some zinc finger proteins bind to a three base triplet in the target strand and a fourth base on the nontarget strand. The fourth base is complementary to the base immediately 3' of the three base subsite.

The structure of the Zif268-DNA complex also suggested that the DNA sequence specificity of a zinc finger protein might be altered by making amino acid substitutions at the four helix positions (-1, 2, 3 and 6) on each of the zinc finger recognition helices. Phage display experiments using zinc finger combinatorial libraries to test this observation were published in a series of papers in 1994 (Rebar et al., *Science* 263, 671-673 (1994); Jamieson et al., *Biochemistry* 33, 5689-5695 (1994); Choo et al, *PNAS* 91, 11163-11167 (1994)). Combinatorial libraries were constructed with randomized side-chains in either the first or middle finger of Zif268 and then used to select for an altered Zif268 binding site in which the appropriate DNA sub-site was replaced by an altered DNA triplet. Further, correlation between the nature of introduced mutations and the resulting alteration in binding specificity gave rise to a partial set of substitution rules for design of ZFPs with altered binding specificity.

Greisman & Pabo, *Science* 275, 657-661 (1997) discuss an elaboration of the phage display method in which each finger of a Zif268 was successively randomized and selected for binding to a new triplet sequence. This paper reported selection of ZFPs for a nuclear hormone response element, a p53 target site and a TATA box sequence.

A number of papers have reported attempts to produce ZFPs to modulate particular target sites. For example, Choo et al., *Nature* 372, 645 (1994), report an attempt to design a ZFP that would repress expression of a brc-abl oncogene. The target segment to which the ZFPs would bind was a nine base sequence 5'GCA GAA3' GCC chosen to overlap the junction created by a specific oncogenic translocation fusing the genes encoding brc and abl. The intention was that a ZFP specific to this target site would bind to the oncogene without binding to abl or brc component genes. The authors used phage display to screen a mini-library of variant ZFPs for binding to this target segment. A variant ZFP thus isolated was then reported to repress expression of a stably transfected brc-able construct in a cell line.

Pomerantz et al., *Science* 267, 93-96 (1995) reported an attempt to design a n vel DNA binding protein by fusing two fingers from Zif268 with a homeodomain from Oct-1. The hybrid protein was then fused with a transcriptional activator for

expression as a chimeric protein. The chimeric protein was reported to bind a target site representing a hybrid of the subsites of its two components. The authors then constructed a reporter vector containing a luciferase gene operably linked to a promoter and a hybrid site for the chimeric DNA binding protein in proximity to the promoter. The authors reported that their chimeric DNA binding protein could activate expression of the luciferase gene.

Liu et al., *PNAS* 94, 5525-5530 (1997) report forming a composite zinc finger protein by using a peptide spacer to link two component zinc finger proteins each having three fingers. The composite protein was then further linked to transcriptional activation domain. It was reported that the resulting chimeric protein bound to a target site formed from the target segments bound by the two component zinc finger proteins. It was further reported that the chimeric zinc finger protein could activate transcription of a reporter gene when its target site was inserted into a reporter plasmid in proximity to a promoter operably linked to the reporter.

Choo et al., WO 98/53058, WO98/53059, and WO 98/53060 (1998) discuss selection of zinc finger proteins to bind to a target site within the HIV Tat gene. Choo et al. also discuss selection of a zinc finger protein to bind to a target site encompassing a site of a common mutation in the oncogene ras. The target site within ras was thus constrained by the position of the mutation.

None of the above studies provided criteria for systematically evaluating the respective merits of the different potential target sites within a candidate gene. The phage display studies by Rebar et al., supra, Jamieson et al., supra and Choo et al, *PNAS*.(1994) supra, all focused on alterations of the natural Zif268 binding-site, 5'GCG TGG GCGc3', and were not made with reference to a predetermined target gene. Choo et al. Nature (1994), supra's selection of target site was constrained solely by the intent that the site overlap the interface between brc and abl segments and did not involve a comparison of different potential target sites. Likewise, Greisman & Pabo chose certain target sites because of their known regulatory roles and did not consider the relative merits of different potential target segments within a preselected target gene. Similarly, Choo et al. (1998), supra's choice of target site within ras was constrained by the position of a mutation. No criterion is provided for Choo et al. (1998)'s selection of a target site in HIV Tat. Finally, both Pomerantz et al., supra and Liu et al., supra constructed artificial

hybrid target sites for composite zinc fingers and then inserted the target sites into reporter constructs.

SUMMARY OF THE INVENTION

The invention provides methods of selecting a target site within a target sequence for targeting by a zinc finger protein. Some such methods comprise providing a target nucleic acid to be targeted by a zinc finger protein and outputting a target site within the target nucleic acid comprising 5'NNx aNy bNzc3'. Each of (x, a), (y, b) and (z, c) is (N, N) or (G, K) provided at least one of (x, a), (y, b) and (z, c) is (G, K). N and K are IUPAC-IUB ambiguity codes. In some methods, a plurality of segments within the target nucleic acid are selected and a subset of the plurality of segments comprising 5'NNx aNy bNzc3' is output. Typically the target nucleic acid comprises a target gene. In some methods, at least two of (x, a), (y, b) and (z, c) is (G, K) In some methods, all three of (x, a), (y, b) and (z, c) are (G, K). Some methods further comprise identifying a second segment of the gene comprising 5'NNx aNy bNzc3', wherein each of (x, a), (y, b) and (z, c) is (N, N) or (G, K); at least one of (x, a), (y, b) and (z, c) is (G, K). and N and K are IUPAC-IUB ambiguity codes. In some methods, in the second segment at least two of (x, a), (y, b) and (z, c) are (G, K). In some methods, all three of at least one of (x, a), (y, b) and (z, c) are (G, K). In some methods, the first and second segments are separated by fewer than 5 bases in the target site.

Some methods further comprise synthesizing a zinc finger protein comprising first, second and third fingers that bind to the bNz aNy and NNx triplets respectively. In some such methods, the synthesizing step comprises synthesizing a first zinc finger protein comprising three zinc fingers that respectively bind to the NNx aNy and bNz triplets in the target segment and a second three fingers that respectively bind to the NNx aNy and bNz triplets in the second target segment. In some methods, each of the first, second and third fingers is selected or designed independently. In some methods, a finger is designed from a database containing designations of zinc finger proteins, subdesignations of finger components, and nucleic acid sequences bound by the zinc finger proteins. In some methods, a finger is selected by screening variants of a zinc finger binding protein for specific binding to the target site to identify a variant that binds to the target site.

Some methods further comprise contacting a sample containing the target nucleic acid with the zinc finger protein, whereby the zinc finger protein binds to the target site revealing the presence of the target nucleic acid or a particular allelic form thereof. In some methods, a sample containing the target nucleic acid is contacted with the zinc finger protein, whereby the zinc finger protein binds to the target site thereby modulating expression of the target nucleic acid.

In some methods, the target site occurs in a coding region. In some methods, the target site occurs within or proximal to a promoter, enhancer, or transcription start site. In some methods, the target site occurs outside a promoter, regulatory sequence or polymorphic site within the target nucleic acid.

In another aspect, the invention provides alternate methods for selecting a target site within a polynucleotide for targeting by a zinc finger protein. These methods, comprising providing a polynucleotide sequence and selecting a potential target site within the polynucleotide sequence; the potential target site comprising contiguous first, second and third triplets of bases at first, second and third positions in the potential target site. A plurality of subscores are then determined by applying a correspondence regime between triplets and triplet position in a sequence of three contiguous triplets, wherein each triplet has first, second and third corresponding positions, and each combination of triplet and triplet position has a particular subscore. A score is then calculated for the potential target site by combining subscores for the first, second, and third triplets. The selecting, determining and calculating steps are then repeated at least once on a further potential target site comprising first, second and third triplets at first, second and third positions of the further potential target site to determine a further score. Output is then provided of at least one potential target site with its score. In some methods, output is provided of the potential target site with the highest score. In some methods, output is provided of the n potential target sites with the highest scores, and the method further comprises providing user input of a value for n. In some methods, the subscores are combined by forming the product of the subscores. In some methods, the correspondence regime comprises 64 triplets, each having first, second, and third corresponding positions, and 192 subscores.

In some methods, the subscores in the correspondence regime are determined by assigning a first value as the subscore of a subset of triplets and corresponding positions, for each of which there is an existing zinc finger protein that

comprising a finger that specifically binds to the triplet from the same position in the existing zinc finger protein as the corresponding position of the triplet in the correspondence regime; assigning a second value as the subscore of a subset of triplets and corresponding positions, for each of which there is an existing zinc finger protein that comprises a finger that specifically binds to the triplet from a different position in the existing zinc finger protein than the corresponding position of the triplet in the correspondence regime; and assigning a third value as the subscore of a subset of triplets and corresponding positions for which there is no existing zinc protein comprising a finger that specifically binds to the triplet.

In some methods, a context parameter with the subscore of at least one of the first, second and third triplets to give a scaled subscore of the at least one triplet. In some methods the context parameter is combined with the subscore when the target site comprises a base sequence 5'NNGK3', wherein NNG is the at least one triplet.

In another aspect, the invention provides methods of designing a zinc finger protein. Such methods use a database comprising designations for a plurality of zinc finger proteins, each protein comprising at least first, second and third fingers, and subdesignations for each of the three fingers of each of the zinc finger proteins; a corresponding nucleic acid sequence for each zinc finger protein, each sequence comprising at least first, second and third triplets specifically bound by the at least first, second and third fingers respectively in each zinc finger protein, the first, second and third triplets being arranged in the nucleic acid sequence (3'-5') in the same respective order as the first, second and third fingers are arranged in the zinc finger protein (N-terminal to C-terminal). A target site is provided for design of a zinc finger protein, the target site comprising continuous first, second and third triplets in a 3'-5' order. For the first, second and third triplet in the target site, first, second and third sets of zinc finger protein(s) in the database are identified, the first set comprising zinc finger protein(s) comprising a finger specifically binding to the first triplet in the target site, the second set comprising zinc finger protein(s) comprising a finger specifically binding to the second triplet in the target site, the third set comprising zinc finger protein(s) comprising a finger specifically binding to the third triplet in the target site. Designations and subdesignations of the zinc finger proteins in the first, second, and third sets identified in st p (c) are then output. Some method further comprise producing a zinc finger protein that binds to the target site comprising a first finger from a zinc finger protein from the

first set, a second finger from a zinc finger protein from the second set, and a third finger from a zinc finger protein from the third set

Some methods further comprises identifying subsets of the first, second and third sets. The subset of the first set comprising zinc finger protein(s) comprising a finger that specifically binds to the first triplet in the target site from the first finger position of a zinc finger protein in the database. The subset of the second set comprising zinc finger protein(s) comprises a finger that specifically binds to the second triplet in the target site from the second finger position in a zinc finger protein in the database; the subset of the third set comprises a zinc finger protein(s) comprising a finger that specifically binds to the third triplet in the target site from a third finger position in a zinc finger protein in the database. Designations and subdesignations of the subset of the first, second and third sets are output. A zinc finger protein comprising a first finger from the first subset, a second finger from the second subset, and a third finger from the third subset is then produced. In some of the above methods of design, the target site is provided by user input. In some methods, the target site is provided by one of the target site selection methods described above.

The invention further provides computer program products for implementing any of the methods described above. One computer program product implements methods for selecting a target site within a polynucleotide for targeting by a zinc finger protein. Such a product comprises (a) code for providing a polynucleotide sequence; (b) code for selecting a potential target site within the polynucleotide sequence; the potential target site comprising first, second and third triplets of bases at first, second and third positions in the potential target site; (c) code for calculating a score for the potential target site from a combination of subscores for the first, second, and third triplets, the subscores being obtained from a correspondence regime between triplets and triplet position, wherein each triplet has first, second and third corresponding positions, and each corresponding triplet and position has a particular subscore; (d) code for repeating steps (b) and (c) at least once on a further potential target site comprising first, second and third triplets at first, second and third positions of the further potential target site to determine a further score; e) code for providing output of at least one of the potential target site with its score; and (f) a computer readable storage medium for holding the codes.

The invention further provides computer systems for implementing any of the methods described above. One such system for selecting a target site within a polynucleotide for targeting by a zinc finger protein, comprises (a) a memory; (b) a system bus; and (c) a processor. The processor is operatively disposed to:(1) provide or receive a polynucleotide sequence; (2) select a potential target site within the polynucleotide sequence; the potential target site comprising first, second and third triplets of bases at first, second and third positions in the potential target site; (3) calculate a score for the potential target site from a combination of subscores for the first, second, and third triplets, the subscores being obtained from a correspondence regime between triplets and triplet position, wherein each triplet has first, second and third corresponding positions, and each corresponding triplet and position has a particular subscore; (4) repeat steps (2) and (3) at least once on a further potential target site comprising first, second and third triplets at first, second and third positions of the further potential target site to determine a further score; (5) provide output of at least one of the potential target site with its score

A further computer program product for producing a zinc finger protein comprises: (a) code for providing a database comprising designations for a plurality of zinc finger proteins, each protein comprising at least first, second and third fingers; subdesignations for each of the three fingers of each of the zinc finger proteins; a corresponding nucleic acid sequence for each zinc finger protein, each sequence comprising at least first, second and third triplets specifically bound by the at least first, second and third fingers respectively in each zinc finger protein, the first, second and third triplets being arranged in the nucleic acid sequence (3'-5') in the same respective order as the first, second and third fingers are arranged in the zinc finger protein (N-terminus to C-terminus); (b) code for providing a target site for design of a zinc finger protein, the target site comprising at least first, second and third triplets; (c) for the first, second and third triplet in the target site, code for identifying first, second and third sets of zinc finger protein(s) in the database, the first set comprising zinc finger protein(s) comprising a finger specifically binding to the first triplet in the target site, the second set comprising a finger specifically binding to the second triplet in the target site, the third set comprising a finger specifically binding to the third triplet in the target site; (d) code for outputting designations and subdesignations of the zinc finger proteins in the first,

second, and third sets identified in step (c) and, (e) a compute readable storage medium for holding the codes.

The invention further provides a system for producing a zinc finger protein. The system comprises (a) a memory; (b) a system bus; and (c) a processor. The processor is operatively disposed to:(1) provide a database comprising designations for a plurality of zinc finger proteins, each protein comprising at least first, second and third fingers, subdesignations for each of the three fingers of each of the zinc finger proteins; a corresponding nucleic acid sequence for each zinc finger protein, each sequence comprising at least first, second and third triplets specifically bound by the at least first, second and third fingers respectively in each zinc finger protein, the first, second and third triplets being arranged in the nucleic acid sequence (3'-5')in the same respective order as the first, second and third fingers are arranged in the zinc finger protein (N-terminus to C-terminus); (2) provide a target site for design of a zinc finger protein, the target site comprising at least first, second and third triplets, (3) for the first, second and third triplet in the target site, identify first, second and third sets of zinc finger protein(s) in the database, the first set comprising zinc finger protein(s) comprising a finger specifically binding to the first triplet in the target site, the second set comprising a finger specifically binding to the second triplet in the target site, the third set comprising a finger specifically binding to the third triplet in the target site; and (4) output designations and subdesignations of the zinc finger proteins in the first, second, and third sets identified in step (3).

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 shows a chart providing data that the presence and number of sub sites in a target site bound by a zinc finger protein correlates with binding affinity.

Fig. 2 shows a three finger zinc finger protein bound to a target site containing three D-able subsites.

Fig. 3 shows the process of assembling a nucleic acid encoding a designed ZFP.

Figs 4 and 5 show computer systems for implementing methods of target site selection and zinc finger protein design

Fig. 6 shows a flow chart of a method for selecting a target site containing a D-able subsite within a target sequence

Fig. 7A shows a flow chart for selecting a target site within a target sequence using a correspondence regime.

Fig. 7B shows a flow chart for designing a ZFP to bind a desired target site using a database.

Fig. 8A is an entity representation diagram of a ZFP database.

Fig. 8B is a representation of a ZFP database.

DEFINITIONS

A zinc finger DNA binding protein is a protein or segment within a larger protein that binds DNA in a sequence-specific manner as a result of stabilization of protein structure through cordination on of zinc ion. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP.

A designed zinc finger protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP designs and binding data. .

A selected zinc finger protein is a protein not found in nature whose production results primarily from an empirical process such as phage display.

The term naturally-occurring is used to describe an object that can be found in nature as distinct from being artificially produced by man. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring. Generally, the term naturally-occurring refers to an object as present in a non-pathological (undiseased) individual, such as would be typical for the species.

A nucleic acid is operably linked when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it increases the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame. However, since enhancers generally function when separated from the promoter by up to several kilobases or more and intronic sequences may be of variabl lengths, some polynucleotide elements may be operably linked but not contiguous.

A specific binding affinity between, for example, a ZFP and a specific target site means a binding affinity of at least 1×10^6 M^{-1}.

The terms "modulating expression" "inhibiting expression" and "activating expression" of a gene refer to the ability of a zinc finger protein to activate or inhibit transcription of a gene. Activation includes prevention of subsequent transcriptional inhibition (i.e., prevention of repression of gene expression) and inhibition includes prevention of subsequent transcriptional activation (i.e., prevention of gene activation). Modulation can be assayed by determining any parameter that is indirectly or directly affected by the expression of the target gene. Such parameters include, e.g., changes in RNA or protein levels, changes in protein activity, changes in product levels, changes in downstream gene expression, changes in reporter gene transcription (luciferase, CAT, beta-galactosidase, GFP (see, e.g., Mistili & Spector, *Nature Biotechnology* 15:961-964 (1997)); changes in signal transduction, phosphorylation and dephosphorylation, receptor-ligand interactions, second messenger concentrations (e.g., cGMP, cAMP, IP3, and Ca2+), cell growth, neovascularization, *in vitro, in vivo, and ex vivo*. Such functional effects can be measured by any means known to those skilled in the art, e.g., measurement of RNA or protein levels, measurement of RNA stability, identification of downstream or reporter gene expression, e.g., via chemiluminescence, fluorescence, colorimetric reactions, antibody binding, inducible markers, ligand binding assays; changes in intracellular second messengers such as cGMP and inositol triphosphate (IP3); changes in intracellular calcium levels; cytokine release, and the like.

A "regulatory domain" refers to a protein or a protein subsequence that has transcriptional modulation activity. Typically, a regulatory domain is covalently or non-covalently linked to a ZFP to modulate transcription. Alternatively, a ZFP can act alone, without a regulatory domain, or with multiple regulatory domains to modulate transcription.

A D-able subsite within a target site has the motif 5'NNGK3'. A target site containing one or more such motifs is sometimes described as a D-able target site. A zinc finger appropriately designed to bind to a D-able subsite is sometimes referred to as a D-able finger. Likewise a zinc finger protein containing at least one finger designed or selected to bind to a target site including at least one D-able subsite is sometimes referred to as a D-able zinc finger protein.

DETAILED DESCRIPTION

I. General

In one aspect, the invention is directed to methods of selecting appropriate segments within a preselected target gene for design of a zinc finger protein intended for use in modulating or detecting the gene. The size of a potential target gene can vary widely from around 100 to several 100,000 bp. A zinc finger protein can bind to a small subsequence or target site within such gene. For example, zinc finger proteins containing three fingers typically bind to nine or ten bases of a target gene. The invention provides criteria and methods for selecting optimum subsequence(s) from a target gene for targeting by a zinc finger protein.

Some of the methods of target site selection seek to identify one or more target segments having a DNA motif containing one or more so-called D-able subsites. A D-able subsite is defined by a characteristic DNA sequence formula as discussed in detail below. A zinc finger protein is able to bind such a motif in a manner such that at least one component finger of the zinc finger protein contacts an additional base outside the three base subsite usually bound by a finger. If two D-able sites are present in the target segment, then two component fingers of a zinc finger protein can each bind to four bases of the target site. If three D-able subsites are present in the target segment, then three component fingers of zinc finger protein can each bind to four bases in the target site. In general zinc finger proteins binding to target sites containing at least one D-able subsite show higher binding affinity than zinc finger proteins that bind to target segments lacking a D-able subsite. Likewise, zinc finger proteins binding to a target site with two D-able subsites generally show higher binding affinity than zinc finger proteins that bind to a target site with one D-able subsite, and zinc finger proteins with three D-able subsites generally show higher binding affinity than zinc finger proteins that bind to a target site with two D-able subsites. Although an understanding of mechanism is not required for practice of the invention, it is believed that the higher binding affinity results from the additional interactions possible between a zinc finger and four bases in a target segment relative to the interactions possible between a zinc finger and three bases in a target segment. In general, the potential for high affinity binding of target segments with D-able subsites makes them the target sites of choice from within target genes for design of zinc finger proteins because higher binding affinity often results in a greater extent of, and/or greater specificity in, modulation of a target gene.

Other methods of the invention are directed to selection of target segments within target genes by additional or alternative criteria to the D-able subsite. The principal criteria for selection of target segments in such methods are provided in the form of a correspondence regime between different triplets of three bases and the three possible positions of a triplet within a nine-base site (i.e., bases 1-3, 4-6 and 7-9). An exemplary correspondence regime is shown in Table 1. The correspondence regime provides different values for different combinations of triplet and triplet position within a target site. A potential target site within a target gene is evaluated by determining a score for the site by combining subscores for its component triplets obtained from the correspondence regime. The scores of different potential target sites are compared, with a high score indicating desirability of a particular segment as a target site for design of zinc finger binding protein.

In another aspect, the invention provides methods of designing zinc finger proteins that bind to a preselected target site. These methods can, of course, be used following the preselection of target sites according to the procedures and criteria described above. The methods of design use a database containing information about previously characterized zinc finger proteins. This information includes names or other designations of previously characterized zinc finger proteins, the amino acid sequence of their component fingers, and the nucleotide triplets bound by each finger of the proteins. Information in the database is accessed using an algorithm that allows one to select fingers from different previous designs for combination in a novel zinc finger protein having specificity for a chosen target site.

II. Zinc Finger Proteins

Zinc finger proteins are formed from zinc finger components. For example, zinc finger proteins can have one to thirty-seven fingers, commonly having 2, 3, 4, 5 or 6 fingers. A zinc finger protein recognizes and binds to a target site (sometimes referred to as a target segment) that represents a relatively small subsequence within a target gene. Each component finger of a zinc finger protein can bind to a subsite within the target site. The subsite includes a triplet of three contiguous bases all on the same strand (sometimes referred to as the target strand). The subsite may or may not also include a fourth base on the opposite strand that is the complement of the base immediately 3' of the three contiguous bases on the target strand. In many zinc finger

proteins, a zinc finger binds to its triplet subsite substantially independently of other fingers in the same zinc finger protein. Accordingly, the binding specificity of zinc finger protein containing multiple fingers is usually approximately the aggregate of the specificities of its component fingers. For example, if a zinc finger protein is formed from first, second and third fingers that individually bind to triplets XXX, YYY, and ZZZ, the binding specificity of the zinc finger protein is 3'XXX YYY ZZZ5'.

The relative order of fingers in a zinc finger protein from N-terminal to C-terminal determines the relative order of triplets in the 3' to 5' direction in the target. For example, if a zinc finger protein comprises from N-terminal to C-terminal the first, second and third fingers mentioned above, then the zinc finger protein binds to the target segment 3'XXXYYYZZZ5'. If the zinc finger protein comprises the fingers in another order, for example, second finger, first finger, third finger, then the zinc finger protein binds to a target segment comprising a different permutation of triplets, in this example, 3'YYYXXXZZZ5' (see Berg & Shi, *Science* 271, 1081-1086 (1996)). The assessment of binding properties of a zinc finger protein as the aggregate of its component fingers is, however, only approximate, due to context-dependent interactions of multiple fingers binding in the same protein.

Two or more zinc finger proteins can be linked to have a target specificity that is the aggregate of that of the component zinc finger proteins (see e.g., Kim & Pabo, *PNAS* 95, 2812-2817 (1998)). For example, a first zinc finger protein having first, second and third component fingers that respectively bind to XXX, YYY and ZZZ can be linked to a second zinc finger protein having first, second and third component fingers with binding specificities, AAA, BBB and CCC. The binding specificity of the combined first and second proteins is thus 3'XXXYYYZZZ___AAABBBCCC5', where the underline indicates a short intervening region (typically 0-5 bases of any type). In this situation, the target site can be viewed as comprising two target segments separated by an intervening segment.

Linkage can be accomplished using any of the following peptide linkers. T G E K P: (Liu et al., 1997, supra.); (G4S)n (Kim et al., *PNAS* 93, 1156-1160 (1996.); GGRRGGGS; LRQRDGERP; LRQKDGGGSERP; LRQKD(G3S)2 ERP. Alternatively, flexible linkers can be rationally designed using computer program capable of modeling both DNA-binding sites and the peptides themselves or by phage display methods . In a further variation, noncovalent linkage can be achieved by fusing two zinc finger proteins

with domains promoting heterodimer formation of the two zinc finger proteins. For example, one zinc finger protein can be fused with fos and the other with jun (see Barbas et al., WO 95/119431).

Linkage of two zinc finger proteins is advantageous for conferring a unique binding specificity within a mammalian genome. A typical mammalian diploid genome consists of 3×10^9 bp. Assuming that the four nucleotides A, C, G, and T are randomly distributed, a given 9 bp sequence is present ~23,000 times. Thus a ZFP recognizing a 9 bp target with absolute specificity would have the potential to bind to ~23,000 sites within the genome. An 18 bp sequence is present once in 3.4×10^{10} bp, or about once in a random DNA sequence whose complexity is ten times that of a mammalian genome.

A component finger of zinc finger protein typically contains about 30 amino acids and has the following motif (N-C) :

Cys-$(X)_{2-4}$-Cys-X.X.X.X.X.X.X.X.X.X.X-**His**-$(X)_{3-5}$-His

-1 1 2 3 4 5 6 7

The two invariant histidine residues and two invariant cysteine residues in a single beta turn are co-ordinated through zinc (see, e.g., Berg & Shi, *Science* 271, 1081-1085 (1996)). The above motif shows a numbering convention that is standard in the field for the region of a zinc finger conferring binding specificity. The amino acid on the left (N-terminal side) of the first invariant His residues is assigned the number +6, and other amino acids further to the left are assigned successively decreasing numbers. The alpha helix begins at residue 1 and extends to the residue following the second conserved histidine. The entire helix is therefore of variable length, between 11 and 13 residues.

The process of designing or selecting a nonnaturally occurring or variant ZFP typically starts with a natural ZFP as a source of framework residues. The process of design or selection serves to define nonconserved positions (i.e., positions -1 to +6) so as to confer a desired binding specificity. One suitable ZFP is the DNA binding domain of the mouse transcription factor Zif268. The DNA binding domain of this protein has the amino acid sequence:

YACPVESCDRRFSRSDELTRHIRIHTGQKP (F1)
FQCRICMRNFSRSDHLTTHIRTHTGEKP (F2)
FACDICGRKFARSDERKRHTKIHLRQK (F3)
and binds to a target 5' GCG TGG GCG 3'.

Another suitable natural zinc finger protein as a source of framework residues is Sp-1. The Sp-1 sequence used for construction of zinc finger proteins corresponds to amino acids 531 to 624 in the Sp-1 transcription factor. This sequence is 94 amino acids in length. The amino acid sequence of Sp-1 is as follows

PGKKKQHICHIQGCGKVYGKTSHLRAHLRWHTGERP
FMCTWSYCGKRFTRSDELQRHKRTHTGEKK
FACPECPKRFMRSDHLSKHIKTHQNKKG

Sp-1 binds to a target site 5'GGG GCG GGG3'.

An alternate form of Sp-1, an Sp-1 consensus sequence, has the following amino acid sequence:

meklrngsgd
PGKKKQHACPECGKSFSKSSHLRAHQRTHTGERP
YKCPECGKSFSRSDELQRHQRTHTGEKP
YKCPECGKSFSRSDHLSKHQRTHQNKKG (lower case letters are a leader sequence from Shi & Berg, *Chemistry and Biology* 1, 83-89. (1995). The optimal binding sequence for the Sp-1 consensus sequence is 5'GGGGCGGGG3'. Other suitable ZFPs are described below.

There are a number of substitution rules that assist rational design of some zinc finger proteins (see Desjarlais & Berg, *PNAS* 90, 2256-2260 (1993); Choo & Klug, *PNAS* 91, 11163-11167 (1994); Desjarlais & Berg, *PNAS* 89, 7345-7349 (1992); Jamieson et al., supra; Choo et al., WO 98/53057, WO 98/53058; WO 98/53059; WO 98/53060). Many of these rules are supported by site-directed mutagenesis of the three-finger domain of the ubiquitous transcription factor, Sp-1 (Desjarlais and Berg, 1992; 1993) One of these rules is that a 5' G in a DNA triplet can be bound by a zinc finger incorporating arginine at position 6 of the recognition helix. Another substitution rule is that a G in the middle of a subsite can be recognized by including a histidine residue at position 3 of a zinc finger. A further substitution rule is that asparagine can be incorporated to recognize A in the middle of triplet, aspartic acid, glutamic acid, serine or threonine can be incorporated to recognize C in the middle of triplet, and amino acids with small side chains such as alanine can be incorporated to recognize T in the middle of triplet. A further substitution rule is that the 3' base of triplet subsite can be recognized by incorporating the following amino acids at position -1 of the recognition helix: arginine to recognize G, glutamine to recognize A, glutamic acid (or aspartic acid) to

recognize C, and threonine to recognize T. Although these substitution rules are useful in designing zinc finger proteins they do not take into account all possible target sites. Furthermore, the assumption underlying the rules, namely that a particular amino acid in a zinc finger is responsible for binding to a particular base in a subsite is only approximate. Context-dependent interactions between proximate amino acids in a finger or binding of multiple amino acids to a single base or vice versa can cause variation of the binding specificities predicted by the existing substitution rules.

The technique of phage display provides a largely empirical means of generating zinc finger proteins with a desired target specificity (see e.g., Rebar, US 5,789,538; Choo et al., WO 96/06166; Barbas et al., WO 95/19431 and WO 98/543111; Jamieson et al., supra). The method can be used in conjunction with, or as an alternative to rational design. The method involves the generation of diverse libraries of mutagenized zinc finger proteins, followed by the isolation of proteins with desired DNA-binding properties using affinity selection methods. To use this method, the experimenter typically proceeds as follows. First, a gene for a zinc finger protein is mutagenized to introduce diversity into regions important for binding specificity and/or affinity. In a typical application, this is accomplished via randomization of a single finger at positions −1, +2, +3, and +6, and sometimes accessory positions such as +1, +5, +8 and +10. Next, the mutagenized gene is cloned into a phage or phagemid vector as a fusion with gene III of a filamentous phage, which encodes the coat protein pIII . The zinc finger gene is inserted between segments of gene III encoding the membrane export signal peptide and the remainder of pIII , so that the zinc finger protein is expressed as an amino-terminal fusion with pIII or in the mature, processed protein. When using phagemid vectors, the mutagenized zinc finger gene may also be fused to a truncated version of gene III encoding, minimally, the C-terminal region required for assembly of pIII into the phage particle. The resultant vector library is transformed into *E. coli* and used to produce filamentous phage which express variant zinc finger proteins on their surface as fusions with the coat protein pIII. If a phagemid vector is used, then the this step requires superinfection with helper phage. The phage library is then incubated with target DNA site, and affinity selection methods are used to isolate phage which bind target with high affinity from bulk phage. Typically, the DNA target is immobilized on a solid support, which is then washed under conditions sufficient to remove all but the tightest binding phage. After washing, any phag remaining on the support are recovered via elution

under conditions which disrupt zinc finger – DNA binding. Recovered phage are used to infect fresh *E. coli*., which is then amplified and used to produce a new batch of phage particles. Selection and amplification are then repeated as many times as is necessary to enrich the phage pool for tight binders such that these may be identified using sequencing and/or screening methods. Although the method is illustrated for pIII fusions, analogous principles can be used to screen ZFP variants as pVIII fusions.

Zinc finger proteins are often expressed with a heterologous domain as fusion proteins. Common domains for addition to the ZFP include, e.g., transcription factor domains (activators, repressors, co-activators, co-repressors), silencers, oncogenes (e.g., myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos family members etc.); DNA repair enzymes and their associated factors and modifiers; DNA rearrangement enzymes and their associated factors and modifiers; chromatin associated proteins and their modifiers (e.g. kinases, acetylases and deacetylases); and DNA modifying enzymes (e.g., methyltransferases, topoisomerases, helicases, ligases, kinases, phosphatases, polymerases, endonucleases) and their associated factors and modifiers. A preferred domain for fusing with a ZFP when the ZFP is to be used for represssing expression of a target gene is a the KRAB repression domain from the human KOX-1 protein (Thiesen et al., *New Biologist* 2, 363-374 (1990); Margolin et al., *Proc. Natl. Acad. Sci. USA* 91, 4509-4513 (1994); Pengue et al., *Nucl. Acids Res.* 22:2908-2914 (1994); Witzgall et al., *Proc. Natl. Acad. Sci. USA* 91, 4514-4518 (1994). Preferred domains for achieving activation include the HSV VP16 activation domain (see, e.g., Hagmann et al., *J. Virol.* 71, 5952-5962 (1997)) nuclear hormone receptors (see, e.g., Torchia et al., *Curr. Opin. Cell. Biol.* 10:373-383 (1998)); the p65 subunit of nuclear factor kappa B (Bitko & Barik, *J. Virol.* 72:5610-5618 (1998)and Doyle & Hunt, *Neuroreport* 8:2937-2942 (1997)); Liu et al., *Cancer Gene Ther.* 5:3-28 (1998)), or artificial chimeric functional domains such as VP64 (Seifpal et al., *EMBO J.* 11, 4961-4968 (1992)).

An important factor in the administration of polypeptide compounds, such as the ZFPs, is ensuring that the polypeptide has the ability to traverse the plasma membrane of a cell, or the membrane of an intra-cellular compartment such as the nucleus. Cellular membranes are composed of lipid-protein bilayers that are freely permeable to small, nonionic lipophilic compounds and are inherently impermeable to polar compounds, macromolecules, and therapeutic or diagnostic agents. However,

proteins and other compounds such as liposomes have been described, which have the ability to translocate polypeptides such as ZFPs across a cell membrane.

For example, "membrane translocation polypeptides" have amphiphilic or hydrophobic amino acid subsequences that have the ability to act as membrane-translocating carriers. In one embodiment, homeodomain proteins have the ability to translocate across cell membranes. The shortest internalizable peptide of a homeodomain protein, Antennapedia, was found to be the third helix of the protein, from amino acid position 43 to 58 (*see, e.g.*, Prochiantz, *Current Opinion in Neurobiology* 6:629-634 (1996)). Another subsequence, the h (hydrophobic) domain of signal peptides, was found to have similar cell membrane translocation characteristics (*see, e.g.*, Lin *et al., J. Biol. Chem.* 270:1 4255-14258 (1995)).

Examples of peptide sequences which can be linked to a ZFP of the invention, for facilitating uptake of ZFP into cells, include, but are not limited to: an 11 animo acid peptide of the tat protein of HIV; a 20 residue peptide sequence which corresponds to amino acids 84-103 of the p16 protein (*see* Fahraeus *et al., Current Biology* 6:84 (1996)); the third helix of the 60-amino acid long homeodomain of Antennapedia (Derossi *et al., J. Biol. Chem.* 269:10444 (1994)); the h region of a signal peptide such as the Kaposi fibroblast growth factor (K-FGF) h region (Lin *et al., supra*); or the VP22 translocation domain from HSV (Elliot & O'Hare, *Cell* 88:223-233 (1997)). Other suitable chemical moieties that provide enhanced cellular uptake may also be chemically linked to ZFPs.

Toxin molecules also have the ability to transport polypeptides across cell membranes. Often, such molecules are composed of at least two parts (called "binary toxins"): a translocation or binding domain or polypeptide and a separate toxin domain or polypeptide. Typically, the translocation domain or polypeptide binds to a cellular receptor, and then the toxin is transported into the cell. Several bacterial toxins, including *Clostridium perfringens* iota toxin, diphtheria toxin (DT), *Pseudomonas* exotoxin A (PE), pertussis toxin (PT), *Bacillus anthracis* toxin, and pertussis adenylate cyclase (CYA), have been used in attempts to deliver peptides to the cell cytosol as internal or amino-terminal fusions (Arora *et al., J. Biol. Chem.*, 268:3334-3341 (1993); Perelle *et al., Infect. Immun.*, 61:5147-5156 (1993); Stenmark *et al., J. Cell Biol.* 113:1025-1032 (1991); Donnelly *et al., PNAS* 90:3530-3534 (1993); Carbonetti *et al., Abstr. Annu. Meet. Am. Soc. Microbiol.* 95:295 (1995); Sebo *et al., Infect. Immun.* 63:3851-3857 (1995); Klimpel

et al., PNAS U.S.A. 89:10277-10281 (1992); and Novak *et al., J. Biol. Chem.* 267:17186-17193 1992)).

Such subsequences can be used to translocate ZFPs across a cell membrane. ZFPs can be conveniently fused to or derivatized with such sequences. Typically, the translocation sequence is provided as part of a fusion protein. Optionally, a linker can be used to link the ZFP and the translocation sequence. Any suitable linker can be used, e.g., a peptide linker.

III. Selection of Target Gene

Zinc finger proteins can be used to modulate the expression of any target polynucleotide sequence. The sequence can be for example, genomic, cDNA or RNA or an expressed sequence tag (EST). Typically, the target polynucleotide includes a gene or a fragment thereof. The term gene is used broadly to include, for example, exonic regions, intronic regions, 5'UTRs, 3' UTRs, 5' flanking sequences, 3' flanking sequences, promoters, enhancers, transcription start sites, ribosome binding sites, regulatory sites, poly-adenylation sites. Target genes can be cellular, viral or from other sources including purely theoretical sequences. Target gene sequences can be obtained from databases, such as GenBank, the published literature or can be obtained *de novo*. Target genes include genes from pathological viruses and microorganisms for which repression of expression can be used to abort infection. Examples of pathogenic viruses include hepatitis (A, B, or C), herpes virus (e.g., VZV, HSV-1, HSV-6, HSV-II, and CMV, Epstein Barr virus), HIV, ebola, adenovirus, influenza virus, flaviviruses, echovirus, rhinovirus, coxsackie virus, cornovirus, respiratory syncytial virus, mumps virus, rotavirus, measles virus, rubella virus, parvovirus, vaccinia virus, HTLV virus, dengue virus, papillomavirus, molluscum virus, poliovirus, rabies virus, JC virus and arboviral encephalitis virus. Some examples of pathogenic bacteria include chlamydia, rickettsial bacteria, mycobacteria, staphylococci, treptocci, pneumonococci, meningococci and conococci, klebsiella, proteus, serratia, pseudomonas, legionella, diphtheria, salmonella, bacilli, cholera, tetanus, botulism, anthrax, plague, leptospirosis, and Lyme disease bacteria.

Target genes also include genes from human or other mammals that contribute to disease. Some such genes are oncogenes, tumor suppressors or growth factors that contribut to cancer. Examples of oncogenes include hMSH2 (Fishel et al.,

Cell 75, 1027-1038 (1993)) and hMLH1 (Papadopoulos et al., *Science* 263, 1625-1628 (1994)). Some examples of growth factors include fibroblast growth factor, platelet-derived growth factor, GM-SCF, VEGF, EPO, Erb-B2, and hGH. Other human genes contribute to disease by rendering a subject susceptible to infection by a microorganism or virus. For example, certain alleles of the gene encoding the CCR5 receptor render a subject susceptible to infection by HIV. Other human genes, such as that encoding amyloid precursor protein or ApoE, contribute to other diseases, such as Alzheimer's disease.

Target genes also include genes of human or other mammals that provide defense mechanisms against diseases due to other sources. For example, tumor repressor genes, provide protection against cancer. Expression of such genes is desirable and zinc finger proteins are used to activate expression.

Target genes also include genes that are normally turned off or expressed at low levels but which through activation can be used to substitute for another defective gene present in some individuals. For example, the fetal hemaglobin genes, which are normally inactive in adult humans, can be activated to substitute for the defective beta-globin gene in individuals with sickle cell anemia.

Target genes also include plant genes for which repression or activation leads to an improvement in plant characteristics, such as improved crop production, disease or herbicide resistance. For example, repression of expression of the FAD2-1 gene results in an advantageous increase in oleic acid and decrease in linoleic and linoleic acids.

IV. Design of Zinc Finger Proteins To Bind D-able Subsites

1. Methods

The invention provides methods that select a target gene, and identify a target site within the gene containing one to six (or more) D-able subsites. A zinc finger protein can then be synthesized that binds to the preselected site. These methods of target site selection are premised, in part, on the present inventors' recognition that the presence of one or more D-able subsites in a target segment confers the potential for higher binding affinity in a zinc finger protein selected or designed to bind to that site relative to zinc finger proteins that bind to target segments lacking D-able subsites. Experimental evidence supporting this insight is provided in Examples 2-9.

A D-able subsite is a region of a target site that allows an appropriately designed single zinc finger to bind to up to four bases rather than up to three of the target site. Such a zinc finger binds to a triplet of bases on one strand of a double-stranded target segment (target strand) and a fourth base on the other strand (see Fig. 2) For a single zinc finger to bind a four base target segment imposes constraints both on the sequence of the target strand and on the amino acid sequence of the zinc finger. The target site within the target strand should include the "D-able" subsite motif 5'NNGK3', in which N and K are conventional IUPAC-IUB ambiguity codes. A zinc finger for binding to such a site should include an arginine residue at position -1 and an aspartic acid, (or less preferably a glutamic acid) at position +2. The arginine residue at position -1 interacts with the G residue in the D-able subsite. The aspartic acid (or glutamic acid) residue at position +2 of the zinc finger interacts with the opposite strand base complementary to the K base in the D-able subsite. It is the interaction between aspartic acid (symbol D) and the opposite strand base (fourth base) that confers the name D-able subsite. As is apparent from the D-able subsite formula, there are two subtypes of D-able subsites: 5'NNGG3' and 5'NNGT3'. For the former subsite, the aspartic acid or glutamic acid at position +2 of a zinc finger interacts with a C in the opposite strand to the D-able subsite. In the latter subsite, the aspartic acid or glutamic acid at position +2 of a zinc finger interacts with an A in the opposite strand to the D-able subsite. In general, NNGG is preferred over NNGT.

In the design of a zinc finger protein with three fingers, a target site should be selected in which at least one finger of the protein, and preferably, two or three fingers have the potential to bind a D-able subsite in a target site. Such can be achieved by selecting a target site from within a larger target gene having the formula

5'NNx aNy bNzc3', wherein

wherein each of the sets (x, a), (y, b) and (z, c) is either (N, N) or (G, K);

at least one of (x, a), (y, b) and (z, c) is (G, K). and

N and K are IUPAC-IUB ambiguity codes.

In other words, at least one of the three sets (x, a), (y, b) and (z, c) is the set (G, K) meaning that the first position of the set is G and the second position is G or T. Those of the three sets (if any) which are not (G, K) are (N, N) meaning that the first position of the set can be occupied by any nucleotide and the second position of the set can be occupied by any nucleotide. As an example, the set (x, a) can be (G, K) and the sets (y, b) and (z, c) can both be (N, N).

In the formula 5'NNx aNy bNzc3', the triplets of NNx aNy and bNz represent the triplets of bases on the target strand bound by the three fingers in a zinc finger protein. The complements of the highlighted bases are the sites of potential fourth base binding on the nontarget strand. If only one of x, y and z is a G, and this G is followed by a K, the target site includes a single D-able subsite. For example, if only x is G and a is K, the site reads NNG KNy bNz w with the D-able subsite highlighted. If both x and y but not z are G and a and b are K, then the target site has two overlapping D-able subsites as follows: 5'NNG *KNG KNz* c3' with one such site being represented in bold and the other in italics. If all three of x, y and z are G and a, b and c are K, then the target segment includes three D-able subsites, as follows 5'NNG *KNG* KNG K3', the D-able subsites being represented by bold, italics and underline.

The methods of the invention thus work by selecting a target gene, and systematically searching within the possible subsequences of the gene for target sites conforming to the formula 5'NNx aNy bNzc3', wherein

wherein each of (x, a), (y, b) and (z, c) is (N, N) or (G, K);

at least one of (x, a), (y, b) and (z, c) is (G, K). and

N and K are IUPAC-IUB ambiguity codes.

In some such methods, every possible subsequence of 10 contiguous bases on either strand of a potential target gene is evaluated to determine whether it conforms to the above formula, and, if so, how many D-able subsites are present. Typically, such a comparison is performed by computer, and a list of target sites conforming to the formula are output. Optionally, such target sites can be output in different subsets according to how many D-able subsites are present.

In a variation, the methods of the invention identify first and second target segments, each independently conforming to the above formula. The two target segments in such methods are constrained to be adjacent or proximate (i.e., within about 0-5 bases) of each other in the target gene. The strategy underlying selection of proximate target segments is to allow the design of a zinc finger protein formed by linkage of two component zinc finger proteins specific for the first and second target segments respectively. These principles can be extended to select target sites to be bound by zinc finger proteins with any number of component fingers. For example, a suitable target site for a nine-finger protein would have three component segments, each conforming to the above formula.

The target sites identified by the above methods can be subject to further evaluation by other criteria or can be used directly for design or selection (if needed) and production of a zinc finger protein specific for such a site. A further criterion for evaluating potential target sites is proximity to particular regions within a gene. If a zinc finger protein is to be used to repress a cellular gene on its own (i.e., without linking the zinc finger protein to a repressing moiety), then the optimal location appears to be at the site of transcription initiation, or within about 50 bp upstream or downstream, or alternatively within an enhancer element to interfere with the formation of the transcription complex (Kim & Pabo, *J. Biol. Chem.* (1997) or compete for an essential enhancer binding protein. If, however, a ZFP is fused to a functional domain such as the KRAB repressor domain or the VP16 activator domain, the choice of location of the binding site is considerably more flexible and can be outside known regulatory regions. For example, a KRAB domain can repress transcription of a promoter up to at least 3-kb from where KRAB is bound. Thus, target sites can be selected that do not include or overlap segments of significance with target genes, such as regulatory sequences, or polymorphic sites. Other criteria for further evaluating target segments include the prior availability of zinc finger proteins binding to such segments or related segments, and/or ease of designing new zinc finger proteins to bind a given target segment. Implementation of such criteria in the selection process is discussed in further detail below.

Once a target segment has been selected, a zinc finger protein that binds to the segment can be provided by a variety of approaches. The simplest approach is to provide a precharacterized zinc finger protein from an existing collection that is already

known to bind to the target site. However, in many instances, such a zinc finger protein does not exist. An alternative approach uses information in a database of existing zinc finger proteins and binding specificities to design new zinc finger proteins. This approach is described in more detail below. A further approach is to design a zinc finger protein based on substitution rules as discussed above. A still further alternative is to select a zinc finger protein with specificity for a given target by an empirical process such as phage display. In some such methods, each component finger of a zinc finger protein is designed or selected independently of other component fingers. For example, each finger can be obtained from a different pre-existing ZFP. or each finger can be subject to separate randomization and selection.

Once a zinc finger protein has been selected, designed, or otherwise provided to a given target segment, the zinc finger protein or the DNA encoding are synthesized. Exemplary methods for synthesizing and expressing DNA encoding zinc proteins are described below. The zinc finger protein or a polynucleotide encoding it can then be used for modulation of expression, or analysis of the target gene containing the target site to which the zinc finger protein binds.

2. D-able Zinc Finger Proteins

A zinc finger protein is described as D-able if it contains a finger that can bind to the fourth base of at least one D-able subsite, that is a polynucleotide sequence 5'NNGK3'. A preferred framework for designing D-able zinc fingers is the human wild type Sp-1 DNA binding domain. The target for the human transcription factor Sp-1 is 5'GGG GCG GGG3', and fingers 1 and 2 of this protein have an R-1 D+2 arrangement. Designed ZFPs can be identical to Sp-1 except in the recognition helix of each of the three fingers, where the sequences are designed to recognize each of the triplets with which they interact. The mouse ZFP Zif268, which binds the site GCG TGG GCG, is also suitable, having the R-1 D+2 arrangement in all three fingers.

Other zinc finger proteins as a source of framework residues for design of zinc finger proteins capable of binding to D-able subsites can be obtained from or derived from ZFPs from several alternative sources. For example, the TTK transcriptional regulatory protein of the fruit fly *Drosophila melanogaster* has been well characterized with regard to both the sequences of its recognition helices and its DNA site. The protein has nly two fingers and binds to a six base target, so finger 2 interacts with the first

DNA triplet and finger 1 recognizes the second triplet of the site. The site is 5'AAG GAT3' with a GG type D-able subsite present at the junction of the first and second triplet, and finger 2 has the R-1 D+2 sequence. Other suitable ZFPs are found in the unicellular eukaryote *Saccharomyces cerevisiae.* The ADR gene product is known to regulate expression of the ADH gene by binding within the ADH promoter. As described above for TTK, the ADR ZFP binding domain has two fingers, and binds to a six base target, TTGGAG. The finger 2 recognition helix has the R-1 D+2 sequence, appropriate for a ZFP binding to a target site with a D-able subsite.

IV. Selection of Target Sites by a Correspondence Regime

The invention further provides additional or alternative methods for selecting a target site from within a target gene. These methods are premised, in part, on the insights that different three-base subsites (triplets) bound by individual fingers have different desirabilities for zinc finger protein design, that these different desirabilities can be expressed as numerical values, and that the numerical values for the three individual triplets comprising a target site can be combined to give an overall score for the target site. The relative merits of different target sites can the be compared from their relative score.

The methods work by providing a polynucleotide sequence typically a gene or cDNA within which one wishes to select a target site for detection or modulation by a ZFP. In practice, one typically provides two sequences for the two strands of a polynucleotide sequence, but for simplicity, the method is illustrated for a single polynucleotide sequence. From within such a polynucleotide sequence, a potential target site of at least 9 bases comprising contiguous first, second and third triplets of bases is selected. The triplets are contiguous in that the first triplet occupies bases 7-9, the second triplet bases 4-6 and the third triplet bases 1-3 of a site, with base 1 in the 5'-3 orientation being designated base 1. This designation of triplets as first, second, and third is arbitrary and could be reversed. However, by designating the first triplet as occupying bases 7-9, the second triplet bases 4-6 and the third triplet bases 1-3, the first, second, and third fingers of a three finger ZFP in an N-C terminal orientation bind to the first, second and third triplets of a target site. Viewed in another manner, the first, second and third fingers in a zinc finger protein order from N terminal to C t rminal are respectively specific for the first, second and third triplets in a target site ordered in the 3'-5' orientation.

A subscore is then determined for each triplet from a correspondence regime between triplets and corresponding positions within a target site. An exemplary correspondence regime is provided in Table 1. The correspondence regime is a matrix providing three values for each triplet at its three possible positions within a nine base target site. The table provides three values for each of the 64 possible triplets. For example, consider a potential target site 5'AAA AAG AAC3'. The AAC triplet occurs in the first position (bases 7-9) of the target site and is assigned a subscore of 1 from Table 1. The AAG triplet occurs in the second position of the target site (bases 4-6) and is assigned a subscore of 8. The AAA triplet occurs in the third position of the target site (bases 1-3) and is assigned a subscore of 8. The subscores of the three triplets in the potential target site are then combined, e.g., by multiplication or addition or some other function. For example, multiplication of the three triplet subscores gives a combined score of $1 \times 8 \times 8 = 64$.

The process is then repeated for a second potential target site. Subscores are determined for each of the three component triplets of the second potential target site, and a combined score is calculated for the second potential target site. The process can then be repeated for further potential target sites. Optionally, the process can be repeated for every possible contiguous subsequence of at least 9 bases in either strand of a target gene of interest. When scores of all potential target sites of interest have been determined the scores are compared. In general a high score indicates desirability of a target site for design of a ZFP. One or more of the target sites identified with high scores can be outputted together with the score.

The designation of values in the correspondence regime can reflect any criteria that make one triplet subsite more desirable than another for zinc finger protein design or selection. The values in the exemplary correspondence regime of Table 1 reflect availability of previously characterized ZFPs known to bind a given nucleotide triplet. If for a given triplet in a given position of a target site, there exist one or more previously characterized ZFPs that specifically bind to a target segment including the triplet at the given position, then the combination of the triplet and given position is assigned a score of 10. If for a given triplet at a given position, there are no previously characterized ZFPs that specifically bind a target site including the triplet at the given position, but there are one or more previously characterized ZFPs that specifically bind to the triplet at a different position, then the triplet is assigned a score of 8. If for a given

triplet and a given position, there are no previously characterized ZFPs that bind the triplet either at the given position or another position, the triplet and position are assigned a value 1.

The values 10, 8 and 1 are only illustrative, and other values could be used. Furthermore, a more sophisticated assignment of values can be used which also takes into account different binding affinities, specificities and presence of D-able sites, among other factors. In such a scheme, combinations of triplets and positions for which prior ZFPs exist with strong binding affinities are typically given higher values than combinations of triplet and positions for which there are prior ZFPs with lower binding affinities.

The selection of potential target sites within a larger sequence and calculation of scores is typically performed by a suitably programmed computer, which outputs one or more potential target site(s) with their score(s). Optionally, user input can be provided to such a computer to specify how many potential target sites should be output. For example, the user can elect to have n potential target sites with the highest scores output, where n is at the discretion of the user. The user can also specify a threshold score, which must be equaled or exceeded for a potential target site to be output.

In a variation of the above method, a potential target site can be evaluated based both on values in a correspondence table and on the presence of one or more D-able subsites. Such is achieved by user input of a context parameter to provide a scaled score for one or more combinations of triplet and a particular position, if the context of the triplet indicates presence of a D-able subsite. For example, a triplet 5'NNG3' followed by an A does not provide a D-able subsite. However, 5'NNG3' followed by a K does provide a D-able site. The user can elect to input a context parameter that increases the value of the subscore for the 5'NNG3' triplet when 5'NNG3' is followed by a K. The scaled subscore for this triplet is then combined with subscores or scaled subscores for other triplets to give an overall score for a potential target site.

In a further variation, a computer performing the above analysis is programmed to output certain target segments receiving high scores in pairs determined by their physical proximity to each other. Paired target segments both of which receive high scores that occur within about five bases of each other are appropriate targets for the

design of six-finger zinc proteins formed by linkage of two component zinc finger proteins each having three fingers.

Potential target sites identified by the above methods can be subject to further evaluation or can be used directly for design or selection (if needed) and production of zinc finger proteins. Zinc finger proteins can be designed and synthesized to such target sites using the same methods described for potential target segments containing D-able subsites described above.

V. Database design of ZFPs

The invention provides methods for design of ZFPs to a preselected target site. These methods are suitable for use in conjunction with the methods of target site selection described above, or by other methods of target site selection.

In designing a new ZFP, it is generally advantageous to make use of information inherent in precharacterized ZFPs and their target sites thereby minimizing the need for *de novo* design or selection. As with target site selection, several factors are involved in this process. Design is facilitated when, for each triplet subsite in a target site, fingers are not only available in existing ZFPs, but such fingers also contact their respective triplet subsites from the same location in the existing proteins as in the proposed design. For example, consider three existing pairs of ZFP and target site: 5'GCG TGG GAC3', bound by a ZFP with fingers F1-F2-F3 (where F3 interacts with GCG, F2 with TGG, and F1 with GAC), 5'AAG GAG GTG3', bound by a ZFP with fingers F4-F5-F6, and 5'CCG TGA GCA3', bound by a ZFP with fingers F7-F8-F9, and a target site 5'GCG GAG GCA3' for which a ZFP is to be designed. In this situation, the novel protein F7-F5-F3 binds to 5'GCG GAG GCA3' with each finger in the novel protein occurring in the same relative position in the novel protein as it did in the database proteins from which it was obtained. This design is advantageous because the analogous environment of each finger in the novel ZFP with that of its previous ZFP means that the finger is likely to bind with similar specificity and affinity in the novel ZFP as in the parent. Thus, the general rule that the binding characteristics of a zinc finger protein are the aggregate of its component fingers is likely to hold.

Novel zinc finger proteins can also be designed from component fingers that are available in existing proteins, but not at the same positions as in the protein to be designed. For example, using the set f existing ZFP-site pairs described above, the

protein F3-F7-F5 can be designed to bind sequence 5'GAG GCA GCG3'. In the novel protein, the fingers occupy different positions than in their respective parental proteins. Although to an approximation a given finger retains its triplet specificity and affinity irrespective of which position it occupies in a ZFP, in practice, contextual effects are more likely to cause changes in specificity and/or affinity of a finger for its triplet subsite when the finger occupies different positions in different zinc finger proteins. Therefore, although ZFPs formed from component fingers occupying different positions than in previously characterized ZFPs typically still bind to the site, the specificity or affinity is sometimes different (typically lower) than expected.

Finally, for preselected target sites including a triplet for which no preexisting finger is available, completely novel fingers can be designed or selected using rules-based approaches or phage display.

The invention provides methods of systematically using a database containing information about existing ZFPs in the design of new ZFPs for a preselected target site according to the principles described above. The organization of a typical database is shown in Table 9. The database typically includes designations for each of a collection of precharacterized ZFPs. The ZFPs can be natural ZFPs or variant ZFPs. The designation can be, for example, the name or a symbol representing each ZFP. The database also includes subdesignations for each of the fingers in a ZFP. Typically, the subdesignations are in the form of amino acid residues occupying selected positions in a finger or fingers. For example, in Table. 9 the subdesignations are the amino acids occupying positions -1 through +6 according to conventional numbering. The database further includes a target nucleic acid segment bound by each zinc finger protein. The nucleic acid segment usually includes three triplets of three bases. The three triplets of bases can be included joined as one sequence or as separate sequences. If bases in a nine base target site are numbered consecutively from the 5' end, a first triplet occupies bases 7-9, a second triplet occupies bases 4-6 and a third triplet occupies bases 1-3. According to this designation of triplet position within a target segment, the first finger of a zinc finger protein (i.e., closest to N-terminus) binds to the first triplet, the second finger to the second triplet, and the third finger to the third triplet. The database can also include additional information such as the binding affinity or dissociation constant of a ZFP for its target site , although such is not essential.

A target site is provided for design of a zinc finger protein using the database. In some methods, the target site is provided by user input. In other methods, the target site is provided as output from any of the methods of target site selection described above. The target site typically comprises at least 9 bases forming at least three triplets. The three component triplets are designated first, second and third triplets respectively occupying bases 7-9, 4-6 and 1-3 of the target site, with the 5' base being assigned as base 1. For the first triplet in the target site, the computer searches the database for a zinc finger protein(s) containing fingers that bind to the triplet. The computer stores records relating to the zinc finger protein(s) thereby identified, and their finger(s) that bind to the first triplet. Optionally, the computer distinguishes between zinc finger proteins containing a finger that binds to the first triplet of the target site at the first finger position and in other positions. If so, the computer stores the two subsets of zinc finger protein(s) as separate records. The process is then repeated for the second triplet in the target site. The computer identifies zinc finger protein(s) containing a finger that specifically binds to the second triplet. Optionally, the computer distinguishes between zinc finger(s) that bind the second triplet from the second position of an existing zinc finger protein or at a different position. Finally, the computer identifies zinc finger protein(s) containing a finger that specifically binds to the third triplet of the target site. Optionally, the computer distinguishes between zinc finger(s) that bind the third triplet from the third position of an existing zinc finger protein or from another position. After searching for ZFPs that bind to each of the first, second and third triplets in the target segment, the computer outputs designations for the ZFPs that have been identified and subdesignations of the fingers that bind to the first, second and third triplets. Optionally, the computer provides separate output of a subset of ZFPs that bind the first triplet from the first finger position, and a subset of ZFPs that bind the first triplet from other positions; and corresponding subsets of ZFPs that bind the second triplet from the second finger position and from other positions, and of ZFPs that bind the third triplet from the third finger position and from other positions.

The information output by the computer can be used in the design and synthesis of novel zinc finger proteins that bind to a preselected target. For example, if the output includes a ZFP1 with a finger X that binds the first triplet of the target, ZFP2 that includes a finger Y that binds to the second triplet of the target, and ZFP3 that includes a finger Z that binds to the third triplet of th target, a novel ZFP can be

synthesized comprising the fingers XYZ in that order (N-terminal to C-terminal). If the computer outputs multiple different zinc finger proteins that contain multiple different fingers that bind to a given triplet, the user can select between the fingers depending on whether a finger binds to a particular triplet position from the same position in the database protein as in the ZFP to be designed. For example, a ZFP1 containing fingers XYZ, in which X binds to a first triplet in a target site is generally preferred to a ZFP2 containing fingers ABC, in which finger C binds to the first triplet in a target site. Thus one would typically use finger X rather than C to occupy the first finger position in a ZFP designed to bind the target segment. Often the computer program identifies two ZFPs, each containing a finger that binds a particular triplet, and in each ZFP, the finger occupies the same position in the database protein from which it derives as in the intended design ZFP. In such cases, one often chooses between the two fingers based on the binding affinity for their respective targets, with higher binding affinity being preferred. Optionally, the computer also provides output of proposed amino acid substitutions to one or more fingers for the corresponding triplet(s) bound by the finger(s).

Although database analysis is primarily illustrated for precharacterized zinc finger proteins having three fingers, such databases can alternatively or additionally store information concerning zinc finger proteins with fewer or greater numbers of fingers. Likewise, such databases can be used in the design of zinc finger proteins having fewer or greater than three fingers. For example, some databases of the invention store information concerning ZFPs with only two fingers as well as or instead of information concerning ZFPs with three fingers. ZFPs with only two fingers have corresponding target sites with only two triplets. The information relating to two-finger ZFPs can be used in the design of three-finger ZFPs that bind to nine base target sites in essentially the same manner described above. However, there is no exact correspondence between the relative positions of two fingers in a two-finger protein with the relative positions of three fingers in a three-finger zinc finger protein. This issue can be addressed in two ways. First, all fingers in a two-finger protein can be effectively treated as occupying different positions than fingers in a three-finger protein. Accordingly, if a two finger protein contains a finger that binds to a given triplet, the computer outputs this information and indicates that th finger does not occur at the same position in the databas two-finger protein as in the three-finger protein to be designed. Alternatively, the first (N-terminal)

finger in a two-finger protein can be considered the equivalent of either the first or second finger in a three-finger protein. The second finger in a two-finger protein can be considered the equivalent of either the second or third finger in a three-finger protein. Accordingly, if the computer identifies a two finger protein with a first (N-terminal) finger binding to a first triplet in a target site for which a zinc finger protein is to be designed, the computer can output that the two finger protein supplies an appropriate finger and at the same position in the database protein as in the three finger protein to be designed.

VII. Production of ZFPs

ZFP polypeptides and nucleic acids encoding the same can be made using routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel et al., eds., 1994)). In addition, nucleic acids less than about 100 bases can be custom ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company (mcrc@oligos.com), The Great American Gene Company (http://www.genco.com), ExpressGen Inc. (www.expressgen.com), Operon Technologies Inc. (Alameda, CA). Similarly, peptides can be custom ordered from any of a variety of sources, such as PeptidoGenic (pkim@ccnet.com), HTI Bio-products, inc. (http://www.htibio.com), BMA Biomedicals Ltd (U.K.), Bio.Synthesis, Inc.

Oligonucleotides can be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage & Caruthers, *Tetrahedron Letts.* 22:1859-1862 (1981), using an automated synthesizer, as described in Van Devanter et al., *Nucleic Acids Res.* 12:6159-6168 (1984). Purification of oligonucleotides is by either denaturing polyacrylamide gel electrophoresis or by reverse phase HPLC. The sequence of the cloned genes and synthetic oligonucleotides can be verified after cloning using, e.g., the chain termination method for sequencing double-stranded templates of Wallace et al., *Gene* 16:21-26 (1981).

Two alternative methods are typically used to create the coding sequences required to express newly designed DNA-binding peptides. One protocol is a PCR-based assembly procedure that utilizes six overlapping oligonucleotides (Fig. 3). Three

oligonucleotides (oligos 1, 3, and 5 in Figure 3) correspond to "universal" sequences that encode portions of the DNA-binding domain between the recognition helices. These oligonucleotides typically remain constant for all zinc finger constructs. The other three "specific" oligonucleotides (oligos 2, 4, and 6 in Fig. 3) are designed to encode the recognition helices. These oligonucleotides contain substitutions primarily at positions -1, 2, 3 and 6 on the recognition helices making them specific for each of the different DNA-binding domains.

The PCR synthesis is carried out in two steps. First, a double stranded DNA template is created by combining the six oligonucleotides (three universal, three specific) in a four cycle PCR reaction with a low temperature annealing step, thereby annealing the oligonucleotides to form a DNA "scaffold." The gaps in the scaffold are filled in by high-fidelity thermostable polymerase, the combination of Taq and Pfu polymerases also suffices. In the second phase of construction, the zinc finger template is amplified by external primers designed to incorporate restriction sites at either end for cloning into a shuttle vector or directly into an expression vector.

An alternative method of cloning the newly designed DNA-binding proteins relies on annealing complementary oligonucleotides encoding the specific regions of the desired ZFP. This particular application requires that the oligonucleotides be phosphorylated prior to the final ligation step. This is usually performed before setting up the annealing reactions. In brief, the "universal" oligonucleotides encoding the constant regions of the proteins (oligos 1, 2 and 3 of above) are annealed with their complementary oligonucleotides. Additionally, the "specific" oligonucleotides encoding the finger recognition helices are annealed with their respective complementary oligonucleotides. These complementary oligos are designed to fill in the region which was previously filled in by polymerase in the above-mentioned protocol. The complementary oligos to the common oligos 1 and finger 3 are engineered to leave overhanging sequences specific for the restriction sites used in cloning into the vector of choice in the following step. The second assembly protocol differs from the initial protocol in the following aspects: the "scaffold" encoding the newly designed ZFP is composed entirely of synthetic DNA thereby eliminating the polymerase fill-in step, additionally the fragment to be cloned into the vector does not require amplification. Lastly, the design of leaving sequence-specific overhangs eliminates the need for

restriction enzyme digests of the inserting fragment. Alternatively, changes to ZFP recognition helices can be created usingconventional site-directed mutagenesis methods.

Both assembly methods require that the resulting fragment encoding the newly designed ZFP be ligated into a vector. Ultimately, the ZFP-encoding sequence is cloned into an expression vector. Expression vectors that are commonly utilized include, but are not limited to, a modified pMAL-c2 bacterial expression vector (New England BioLabs or an eukaryotic expression vector, pcDNA (Promega). The final constructs are verified by sequence analysis.

Any suitable method of protein purification known to those of skill in the art can be used to purify ZFPs of the invention (see, Ausubel, supra, Sambrook, supra). In addition, any suitable host can be used for expression, e.g., bacterial cells, insect cells, yeast cells, mammalian cells, and the like.

Expression of a zinc finger protein fused to a maltose binding protein (MBP-ZFP) in bacterial strain JM109 allows for straightforward purification through an amylose column (NEB). High expression levels of the zinc finger chimeric protein can be obtained by induction with IPTG since the MBP-ZFP fusion in the pMal-c2 expression plasmid is under the control of the tac promoter (NEB). Bacteria containing the MBP-ZFP fusion plasmids are inoculated in to 2xYT medium containing 10μM ZnCl2, 0.02% glucose, plus 50 μg/ml ampicillin and shaken at 37°C. At mid-exponential growth IPTG is added to 0.3 mMand the cultures are allowed to shake. After 3 hours the bacteria are harvested by centrifugation, disrupted by sonication or by passage through a french pressure cell or through the use of lysozyme, and insoluble material is removed by centrifugation. The MBP-ZFP proteins are captured on an amylose-bound resin, washed extensively with buffer containing 20 mM Tris-HCl (pH 7.5), 200 mM NaCl, 5 mM DTT and 50 μM ZnCl2 , then eluted with maltose in essentially the same buffer (purification is based on a standard protocol from NEB). Purified proteins are quantitated and stored for biochemical analysis.

The dissociation constants of the purified proteins, e.g., Kd, are typically characterized via electrophoretic mobility shift assays (EMSA) (Buratowski & Chodosh, in *Current Protocols in Molecular Biology* pp. 12.2.1-12.2.7 (Ausubel ed., 1996)). Affinity is measured by titrating purified protein against a fixed amount of labeled double-stranded oligonucleotide target. The target typically comprises the natural binding site sequence flanked by the 3 bp found in the natural sequence and additional,

constant flanking sequences. The natural binding site is typically 9 bp for a three-finger protein and 2 x 9 bp + intervening bases for a six finger ZFP. The annealed oligonucleotide targets possess a 1 base 5' overhang which allows for efficient labeling of the target with T4 phage polynucleotide kinase. For the assay the target is added at a concentration of 1 nM or lower (the actual concentration is kept at least 10-fold lower than the than the expected dissociation constant), purified ZFPs are added at various concentrations, and the reaction is allowed to equilibrate for at least 45 min. In addition the reaction mixture also contains 10 mM Tris (pH 7.5), 100 mM KCl, 1 mM MgCl2, 0.1 mM ZnCl2, 5 mM DTT, 10% glycerol, 0.02% BSA. (NB: in earlier assays poly d(IC) was also added at 10-100 μg/μl.)

The equilibrated reactions are loaded onto a 10% polyacrylamide gel, which has been pre-run for 45 min in Tris/glycine buffer, then bound and unbound labeled target is resolved by electrophoresis at 150V. (alternatively, 10-20% gradient Tris-HCl gels, containing a 4% polyacrylamide stacker, can be used) The dried gels are visualized by autoradiography or phosphorimaging and the apparent Kd is determined by calculating the protein concentration that gives half-maximal binding.

The assays can also include determining active fractions in the protein preparations. Active fractions are determined by stoichiometric gel shifts where proteins are titrated against a high concentration of target DNA. Titrations are done at 100, 50, and 25% of target (usually at micromolar levels).

IX. Applications of Designed ZFPs

ZPFs that bind to a particular target gene, and the nucleic acids encoding them, can be used for a variety of applications. These applications include therapeutic methods in which a ZFP or a nucleic acid encoding it is administered to a subject and used to modulate the expression of a target gene within the subject (see copending application Townsend & Townsend & Crew Attorney Docket 019496-002200, filed January 12, 1998). The modulation can be in the form of repression, for example, when the target gene resides in a pathological infecting microrganisms, or in an endogenous gene of the patient, such as an oncogene or viral receptor, that is contributing to a disease state. Alternatively, the modulation can be in the form of activation when activation of expression or increased expression of an endogenous c llular gene can ameliorate a diseased state. For such applications, ZFPs, or more typically, nucleic acids encoding

them are formulated with a pharmaceutically acceptable carrier as a pharmaceutical composition.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. (*see, e.g., Remington's Pharmaceutical Sciences*, 17th ed. 1985)). The ZFPs, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like. Formulations suitable for parenteral administration, such as, for example, by intravenous, intramuscular, intradermal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. Compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials. Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described.

The dose administered to a patient should be sufficient to effect a beneficial therapeutic response in the patient over time. The dose is determined by the efficacy and K_d of the particular ZFP employed, the target cell, and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose also is determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular patient

In other applications, ZFPs are used in diagnostic methods for sequence specific detection of target nucleic acid in a sample. For example, ZFPs can be used to detect variant alleles associated with a disease or phenotype in patient samples. As an example, ZFPs can be used to detect the presence of particular mRNA species or cDNA in a complex mixtures of mRNAs or cDNAs. As a further example, ZFPs can be used to quantify copy number of a gene in a sample. For example, detection of loss of one copy

of a p53 gene in a clinical sample is an indicator of susceptibility to cancer. In a further example, ZFPs are used to detect the presence of pathological microorganisms in clinical samples. This is achieved by using one or more ZFPs specific to genes within the microorganism to be detected. A suitable format for performing diagnostic assays employs ZFPs linked to a domain that allows immobilization of the ZFP on an ELISA plate. The immobilized ZFP is contacted with a sample suspected of containing a target nucleic acid under conditions in which binding can occur. Typically, nucleic acids in the sample are labeled (e.g., in the course of PCR amplification). Alternatively, unlabelled probes can be detected using a second labelled probe. After washing, bound-labelled nucleic acids are detected.

ZFPs also can be used for assays to determine the phenotype and function of gene expression. Current methodologies for determination of gene function rely primarily upon either overexpression or removing (knocking out completely) the gene of interest from its natural biological setting and observing the effects. The phenotypic effects observed indicate the role of the gene in the biological system.

One advantage of ZFP-mediated regulation of a gene relative to conventional knockout analysis is that expression of the ZFP can be placed under small molecule control. By controlling expression levels of the ZFPs, one can in turn control the expression levels of a gene regulated by the ZFP to determine what degree of repression or stimulation of expression is required to achieve a given phenotypic or biochemical effect. This approach has particular value for drug development. By putting the ZFP under small molecule control, problems of embryonic lethality and developmental compensation can be avoided by switching on the ZFP repressor at a later stage in mouse development and observing the effects in the adult animal. Transgenic mice having target genes regulated by a ZFP can be produced by integration of the nucleic acid encoding the ZFP at any site *in trans* to the target gene. Accordingly, homologous recombination is not required for integration of the nucleic acid. Further, because the ZFP is trans-dominant, only one chromosomal copy is needed and therefore functional knock-out animals can be produced without backcrossing.

X. Computer Systems and Programs

Fig. 4 depicts a representative computer system suitable for implementing the present invention. Fig. 4 shows basic subsystems of a computer system 10 suitable for use with the present invention. In Fig. 4, computer system 10 includes a bus 12 which interconnects major subsystems such as a central processor 14, a system memory 16, an input/output controller 18, an external device such as a printer 20 via a parallel port 22, a display screen 24 via a display adapter 26, a serial port 28, a keyboard 30, a fixed disk drive 32 and a floppy disk drive 33 operative to receive a floppy disk 33A. Many other devices can be connected such as a scanner 60 (not shown) via I/O controller 18, a mouse 36 connected to serial port 28 or a network interface 40. Many other devices or subsystems (not shown) may be connected in a similar manner. Also, it is not necessary for all of the devices shown in Fig. 4 to be present to practice the present invention, as discussed below. The devices and subsystems may be interconnected in different ways from that shown in Fig. 4. The operation of a computer system such as that shown in Fig. 4 is readily known in the art and is not discussed in detail in the present application. Source code to implement the present invention may be operably disposed in system memory 16 or stored on storage media such as a fixed disk 32 or a floppy disk 33A.

Fig. 5 is an illustration of representative computer system 10 of Fig. 4 suitable for embodying the methods of the present invention. Fig. 5 depicts but one example of many possible computer types or configurations capable of being used with the present invention. Fig. 5 shows computer system 10 including display screen 24, cabinet 20, keyboard 30, a scanner 60, and mouse 36. Mouse 36 and keyboard 30 illustrate "user input devices." Other examples of user input devices are a touch screen, light pen, track ball, data glove, etc.

In a preferred embodiment, System 10 includes a Pentium® class based computer, running Windows® Version 3.1, Windows95® or Windows98® operating system by Microsoft Corporation. However, the method is easily adapted to other operating systems without departing from the scope of the present invention.

Mouse 36 may have one or more buttons such as buttons 37. Cabinet 20 houses familiar computer components such as disk drive 33, a processor, storage means, etc. As used in this specification "storage means" includes any storage device used in connection with a computer system such as disk drives, magnetic tape, solid state memory, bubble memory, etc. Cabinet 20 may include additional hardware such as

input/output (I/O) interface 18 for connecting computer system 10 to external devices such as a scanner 60, external storage, other computers or additional peripherals. Fig. 5 is representative of but one type of system for embodying the present invention. Many other system types and configurations are suitable for use in conjunction with the present invention.

Fig. 6 depicts a flowchart 301 of simplified steps in a representative embodiment for selecting a target site containing a D-able subsite within a target sequence for targeting by a zinc finger protein. In a step 302, a target sequence to be targeted by a zinc finger protein is provided. Then, in a step 303, a potential target site within the target sequence is selected for evaluation. In a decisional step 304, the potential target site is evaluated to determine whether it contains a D-able subsite. Such a target site conforms to the formula

5'NNx aNy bNzc3', wherein

wherein each of (x, a), (y, b) and (z, c) is (N, N) or (G, K);

at least one of (x, a), (y, b) and (z, c) is (G, K) and

N and K are IUPAC-IUB ambiguity codes.

If the potential target site does contain a D-able subsite, the potential target site is stored as a record in 205. The methods continues with a further decisional step 306. If evaluation of further potential target sites is required by the user, a further iteration of the method is performed starting from 303. If sufficient potential target sites have already been evaluated, records of target sites stored in step 305 are then ouput in step 307.

Fig. 7A depicts a flowchart of simplified steps in another representative embodiment for selecting a target site within a polynucleotide for targeting by a zinc finger protein. In a step 402, a polynucleotide target sequence is provided for analysis. Then, in a step 404, a potential target site within the polynucleotide sequence is selected. The potential target site comprises first, second and third triplets of bases at first, second and third positions in the potential target site. Then, in a step 406, a plurality of subscores are determined by applying a correspondence regime between triplets and triplet position, wherein each triplet has first, second and third corresponding positions, and each

corresponding triplet and position is assigned a particular subscore. Next there is an optional decisional step 408 in which the user can elect to scale one or more of the subscores with a scaling factor in step 410. Thereafter in a step 412, a score is determined from the subscores (scaled as appropriate) for the first, second, and third triplets. Then, in a decisional step 414, a check is performed to determine if any further potential target sites are to be examined. If so, then processing continues with step 404. Otherwise, in a step 416, at least one of the potential target sites and its score are provided as output.

Fig. 7B depicts a flowchart of simplified steps in a representative embodiment for producing a zinc finger protein. In a step 450 a database comprising designations for a plurality of zinc finger proteins is provided. Each protein in the database comprises at least first, second and third fingers. The database further comprises subdesignations for each of the three fingers of each of the zinc finger proteins and a corresponding nucleic acid sequence for each zinc finger protein. Each sequence comprises at least first, second and third triplets specifically bound by the at least first, second and third fingers respectively in each zinc finger protein. The first, second and third triplets have an arrangement in the nucleic acid sequence in the same respective order (3'-5') as the first, second and third fingers are arranged in the zinc finger protein (N-terminal to C-terminal).

In a step 452, a target site for design of a zinc finger proteins comprising at least first, second and third triplets is provided. Then, in a step 454, a first set of zinc finger proteins with a finger that binds to the first triplet in the target sequence is identified. There follows an optional step 456 of identifying first and second subsets of the set determined in 454. The first subset comprises zinc finger protein(s)s with a finger that binds the first triplet from the first finger position in the zinc finger protein. The second subset comprises zinc finger protein(s) with a finger that binds the first triplet from other than the first finger position in the zinc finger protein. The method continues at step 458. In this step, a further set of zinc finger proteins is identified, this set comprising a finger that binds to the second triplet in the target site. This step is followed by an optional step 460 of identifying first and second subsets of the set identified in step 458. The first subset comprises zinc finger protein(s) that bind to the second triplet from the second position within a zinc finger protein. The second subset comprises zinc finger protein(s) that bind the second triplet from other than the second position of a zinc finger

protein. The method continues at step 462. In 462, a set of zinc finger proteins is identified comprising a finger that binds to the third triplet of the target site. In an optional step 464, first and second subsets of the set identified in step 462 are identified. The first subset comprises zinc finger protein(s) containing a finger that binds to the third triplet from the third finger position of the zinc finger protein. The second subset comprises zinc finger protein(s) containing a finger that binds to the third triplet from other than the third finger position of the zinc finger protein. The method continues at step 466 in which the sets of zinc finger protein identified in steps 454, 458 and 462 are separately output. There is a further optional step 468 in which the first and second subsets of zinc finger proteins identified in steps 460, 464 and 468 are output.

Fig. 8A is a key to the Entity Representation Diagram (ERD) that will be used to describe the contents of ZFP database. A representative table 502 includes one or more key attributes 504 and one or more non-key attributes 506. Representative table 502 includes one or more records where each record includes fields corresponding to the listed attributes. The contents of the key fields taken together identify an individual record. In the ERD, each table is represented by a rectangle divided by a horizontal line. The fields or attributes above the line are key while the fields or attributes below the line are non-key fields. An identifying relationship 508 signifies that the key attribute of a parent table 510 is also a key attribute of a child table 512. A non-identifying relationship 514 signifies that the key attribute of a parent table 516 is also a non-key attribute of a child table 518. Where (FK) appears in parenthesis, it indicates that an attribute of one table is a key attribute of another table. For both the non-identifying and the identifying relationships, one record in the parent table corresponds to one or more records in the child table.

Fig. 8B depicts a representative ZFP database 550 according to a particular embodiment of the present invention. Database 550 can typically include designations for each of a collection of precharacterized ZFPs. The ZFPs can be natural ZFPs or variant ZFPs. The designation can be, for example, the name or a symbol representing each ZFP. For example, ZFP 552 of database 550 in Fig. 8B is designated "ZFP001." The database 550 also includes subdesignations for each of the fingers in a ZFP, such as subdesignation 554, Finger 1 of ZFP001 552. Typically, the subdesignations are in the form of amino acid residues occupying selected positions in a finger. Further, the ZFPs have subdesignations that are the amino acids occupying positions -1 through +6

according to conventional numbering. The database can further include a target nucleic acid segment bound by each zinc finger protein. The nucleic acid segment usually includes three triplets of three bases. The three triplets of bases can be included joined as one sequence or as separate sequences. If bases in a nine base target site are numbered consecutively from the 5' end, a first triplet occupies bases 7-9, a second triplet occupies bases 4-6 and a third triplet occupies bases 1-3. According to this designation of triplet position within a target segment, the first finger of a zinc finger protein (i.e., closest to N-terminus) binds to the first triplet, the second finger to the second triplet, and the third finger to the third triplet. The database can also include additional information such as the binding affinity or dissociation constant of a ZFP for its target site, although such is not essential. Further database 550 can include other arrangements and relationships among the ZFPs, fingers and nucleic acids than are depicted in Fig. 8B without departing from the scope of the present invention.

Examples

Example 1: SEARCH PROTOCOLS FOR DNA MOTIFS

This Example illustrates how a target segment is selected from a longer gene. The search procedure is implemented using a computer program that allows one to specify one or more DNA sequence motifs in a search protocol. Normal procedure is to input the DNA sequence of a gene or cDNA and then search the sequence multiple times for different motifs, from the most to the least desirable. Thus, of the exemplary protocols listed below one would typically perform protocol 1 first, and if that fails to yield an adequate number of potential target segments, one then tries protocol 2, and so forth.

Protocol 1 searches a target gene for a target site formed from two separate segments, each of 9 or 10 bases. The two segments can be separated by zero to three intervening bases. Each segment includes a D-able subsite of the form NNGG (shown in bold). Each three base subsite within a segment begins with a G. The target sites identified by this analysis can be used directly for ZFP design or can be subject to further analysis, for example, to identify which target segments possess additional D-able subsites. In a target site formed from two segments, each of ten bases, a total of six D-able subsites can be present. All target sites below are shown from 5' to 3' and the nomenclature "0,3" indicates that 0-3 nucleotide of any type may be present.

GNGGNNGNN(N){0,3}GNGGNNGNNN

GNGGNNGNN(N){0,3}GNNGNGGNNN

GNGGNNGNN(N){0,3}GNGGNNGNGG

GNNGNGGNN(N){0,3}GNGGNNGNNN

GNNGNGGNN(N){0,3}GNNGNGGNNN

GNNGNGGNN(N){0,3}GNGGNNGNGG

GNNGNNGNGG(N){0,3}GNGGNNGNNN

GNNGNNGNGG(N){0,3}GNNGNGGNNN

GNNGNNGNGG(N){0,3}GNGGNNGNGG

GNNGNNGNGGNGGNNGNNN

GNNGNNGNGGNNGNGGNNN

GNNGNNGNGGNNGNNGNGG

Protocol 2 is a second procedure for evaluating target sites within a target gene. This procedures again searches for a target site formed from two segments, each of 9 or 10 bases. Each segment contains at least one D-able subsite of the form KNGG. Protocol 2 differs from protocol 1 in that protocol 2 does not require that three base subsites being with a G. Rather in protocol 2, three bases subsites beginning with either a G or T (K in IUBPAC-IUB ambiguity code). Target sites are shown from 5' to 3', and the symbolds "(0,3) and (0,2) indicate intervening segments of 0-3 and 0-2 bases respetively.

KNGGNNKNN(N){0,3}KNGGNNKNNN

KNGGNNKNN(N){0,3}KNNKNGGNNN

KNGGNNKNN(N){0,3}KNNKNNKNGG

KNNKNGGNN(N){0,3}KNGGNNKNNN

KNNKNGGNN(N){0,3}KNNKNGGNNN

KNNKNGGNN(N){0,3} KNNKNNKNGG

KNNKNNKNGG(N){0,2}KNGGNNKNNN

KNNKNNKNGG(N){0,2}KNNKNGGNNN

KNNKNNKNGG(N){0,2}KNNKNNKNGG

KNNKNNKNGGNGGNNKNNN

KNNKNNKNGGNNKNGGNNN

KNNKNNKNGGNNKNNKNGG

Protocol 3 is the same as protocol two except that protocol three selects target sites with either a KNGG or a KNGT D-able subsite. Target sites are shown from 5'-3'.

KNGKNNKNN(N){0,3}KNGKNNKNNN

KNGKNNKNN(N){0,3}KNNKNGKNNN

KNGKNNKNN(N){0,3}KNNKNNKNGK

KNNKNGKNN(N){0,3}KNGKNNKNNN

KNNKNGKNN(N){0,3}KNNKNGKNNN

KNNKNGKNN(N){0,3}KNNKNNKNGK

KNNKNNKNGK(N){0,2}KNGKNNKNNN

KNNKNNKNGK(N){0,2}KNNKNGKNNN

KNNKNNKNGK(N){0,2}KNNKNNKNGK

KNNKNNKNGKNGKNNKNNN

KNNKNNKNGKNNKNGKNNN

KNNKNNKNGKNNKNNKNGK

Protocol 4 is more general than any of the protocols described above, and does not require that target sites contain a D-able subsite. Protocol 4 similar requires two segments, each of 9 bases within 0-3 bases of each other of the form GNN GNN GNN.

Protocol 5 is the same as protocol 4 except that it searches for target sites formed from two target segments of formula 5'KNN KNN KNN3' within 0-3 bases of each other.

Example 2

This example illustrates that zinc finger proteins that bind to target segments including at least one D-able subsite generally bind with higher affinity than zinc finger proteins binding to target segments lacking D-able subsites provided the ZFP has a D residues at position +2. Fifty-three ZFPs, each having three fingers, were selected from a collection without regard to binding affinity or binding to a D-able subsite. The dissociation constants of the selected ZFPs were determined by binding of the ZFPs to a target segment comprising three contiguous nucleotide triplets respectively bound by the three fingers of the ZFP plus at least one flanking base from the target sequence on either side. All ZFPs had the human Sp1 framework. The binding affinities of these 53 ZFPs were arbitrarily divided into 4 groups, listed as Kd values in Table 2.

Table 2

Dissociation Constants (Kd)			
>1,000nM	100-1,000nM	10-99nM	< or =10nM
31	8	11	3

According to this classification only about 25% (14/53) of these proteins had high affinities (Kd less than or equal to 100 nM) for their respective targets. Of these 14 proteins, all had had least one D-able subsite within the target.

Example 3

We searched the sequence of the soybean (Glycine max) FAD2-1 cDNA for paired proximate 9 base target segments using protocols 2 and 3. Five targets segments were chosen, and either one or two ZFPs were designed to bind to each of the targets. The

targets chosen and the Kd values for the respective designed ZFPs are shown in Table 3. D-able subsites are shown in bold. Sequences are shown from 5' to 3'.

Table 3

TARGET NAME	SEQUENCE	PROTEIN NAME	Kd (nM)
FAD 1	**GAG GTA GAG G**	FAD 1A	10
FAD 1	**GAG GTA GAG G**	FAD 1B	10
FAD 2	**GTC GTG TGG A**	FAD 2A	100
FAD 3	**GTT GAG GAA G**	FAD 3A	100
FAD 3	**GTT GAG GAA G**	FAD 3B	100
FAD 4	**GAG GTG GAA G**	FAD 4A	10
FAD 4	**GAG GTG GAA G**	FAD 4B	2
FAD 5	**TAG GTG GTG A**	FAD 5A	10

Of the 8 ZFPs made, all bound with high affinity (Kd less than or equal to 100 nM) to their targets, showing that selecting target with a D-able subsite within a 9bp target allows one to efficiently design a high affinity ZFP. Moreover, all of the ZFPs binding to target sites with two D-able subsites bound more strongly than ZPFs binding to target sites with only one D-able subsite.

Example 4

This example provides further evidence that D-able subsites confer high binding affinity. Fifty-three target segments were identified by protocol 5 listed above, which does not require that a D-able subsite be present in a target site. Fifty-three ZFPs were designed to bind to these respective sites. Thirty three target segments were identified by protocol 3 above, which does require a D-able subsite, and thirty-three ZFPs were designed to bind to these respective sites. Table 4 compares the Kds of ZFPs designed by the different procedures.

Table 4

Search Protocol	Dissociation Constants (Kd) >1,000nM	100-1,000 nM	10-100nM	< or =10nM
#5	31	8	11	3
#3	0	2	15	16

Table 4 shows that 31 of 33 ZFPs designed by protocol 3 have high binding affinity (Kd less than 100 nM). By contrast, only 14 of 56 ZFPs designed by protocol 5 have high binding affinity. These data show that high affinity ZFPs (Kd<100nM) can be designed more efficiently to targets if the search protocol includes D-able subsite criteria than if the search protocol does not require a D-able subsite.

Example 5

The the relationship between the affinity of the ZFP and the presence of one or more D-able subsites in the target was analyzed for about 300 designed ZFPs specific mostly to different target sites. In this and subsequent analyses, only one ZFP was included per target site, this being the ZFP with the highest affinity.

Table 5 and Fig. 1 show the average Kd of different categories of ZFP categorized by number and type of D-able subsites in 9 base target site bound. In Table 4, and later in Tables 6, 7 and 8, s.e.m. is standard error of the mean , and n is number of proteins examined.

Table 5

D-able subsite/ 9 base target	Aver Kd	n
0	828 (± 66)	24
1 GT	46 (± 226)	05
1 GG	138 (± 35)	34
2 GT	100 (± 30)	02
1 GG+1GT	208 (± 198)	04
2 GG	15 (± 6)	22

The 22 ZFPs designed to targets with two GG type D-able subsites have the strongest binding affinity with an average Kd = 15 nM. Of the 50 ZFPs having a Kd < 100 nM, 49 have at least one D-able subsite. The table shows the following conclusion: (1) binding to a target site with one D-able subsite bind more strongly than ZFPs binding to a target site lacking a D-able subsites; (2) ZFPS binding to a target site with two D-able subsites bind more strongly than ZFPs that bind to a target sing with one D-able subsite, and (3) ZFPs with a target site with a GG D-able subsite bind more strongly than ZFPs with a target site with a GT D-able subsite.

Example 6

Another factor affecting binding affinity of designed ZFPs is whether a target site has the form GNN GNN GNN rather than KNN KNN KNN. his example shows that D-able subsites confer high binding affinity even in the context of a GNN GNN GNN motif. For this analysis, we selected a population of 59 ZFPs, each of which binds to a different target site of the form GNN GNN GNN. Table 6 shows the Kd values of designed ZFPs as a function of the presence of D-able subsites with a GNN GNN GNN target.

Table 6

D-able subsites/ 9bp Target	Average Kd	n
0	787 (± 88)	17
1 GG	66 (± 14)	23
2 GG	17 (± 7)	18
1 GG+1GT	5.5 (± 4.5)	2

The presence of a D-able subsite strongly affects binding affinity of a ZFP even when the target fits the GNN GNN GNN motif.

Example 7

This Example provides further evidence that the effect of D-able subsites in conferring increased binding affinity is additive with any effects of G residues in conferring higher binding affinity relative to other residues. For this analysis, we selected 101 zinc finger proteins binding to different target sites from our collection, and classified these target sites by the number of G residues present. The target sites contained from 2-8 G residues in a 9 base sequence. Table 7 shows that in general, the more G residues that are present in a target site, the stronger the binding affinity of the ZFP for that site.

Table 7

# Gs/ 9 base target	Aver Kd, nM (+/-) s.e.m	n
2	>1000	4
3	681 (± 158)	8
4	447 (± 84)	26
5	195 (± 58)	28
6	83 ± (66)	15
7	46 ± (26)	9
8	1 -	1

We analyzed these data further by asking whether the presence or absence of a D-able subsite affected average Kd values of the designed ZFPs. Each category of 9 base target from Table 7 was subdivided into targets containing or not containing D-able subsites. The result of this analysis is shown in Table 8.

Table 8

	Kd, nM D-able site?	
G's/Target	minus	plus
3	809 ± 191	467 ± 273
4	867 ± 87	132 ± 5
5	640 ± 169	98 ± 39
6	>1000	8 ± 66

The table shows that when target sites having the same number of G residues but different numbers of D-able subsites are compared, the sites including D-able subsite(s) confer higher binding affinity. For 9 base target sites having has 4 or more Gs, the average Kd is approximately 100 nM or less if the target has at least one D-able subsite. Particularly notable is the comparison between target sites having 5 G residues. 5 such target sites lacking a D-able subsite had an average Kd of 640 nM. 23 such target sites with two D-able subsites had an average Kd of 98 nM.

Example 10: The ZFP Prediction Module

This example illustrates selection of a target segment within a target gene using a correspondence regime, and use of a database to design a ZFP that binds to the selected target segment. The ZFP Prediction Module facilitates both the site selection and ZFP design processes by taking as input (i) the DNA sequence of interest (ii) various data tables (iii) design parameters and (iv) output parameters, and providing as output a list of potential ZFP target sites in the sequence of interest and a summary of fingers which have been designed to subsites in each target site. This section will describe program inputs, outputs, and scoring protocols for the program. For clarity, the descriptions will be divided into site selection and design functions.

1. Selection of target sites within the DNA region of interest:

Inputs:

1) The target DNA sequence

2) A scores table listing each of the possible three-base pair subsites and scores for its three possible locations in a 9-bp target site is shown in Table 1. The scores table is provided by the user at run-time and may be customized and updated to reflect the user's most current understanding of the DNA-sequence preferences of the zinc finger motif.

3) A 'ZFP data table' which contains target sites, amino acid sequences, and reference data for existing high-affinity ZFPs. This table is required for this portion of the program only if output parameter (ii) is selected below. An example of a ZFP data table is provided in Table 9.

4) An optional context parameter - the "enhancement factor for 'D-able' triplets" - entered by the user at run-time. This parameter multiplies - by the enhancement factor - the score for any 'xxG' subsite flanked by a 3' G or T.

5) Output parameters - supplied by the user - specifying

i) the number of target sites to include in the output

ii) whether the program should specifically highlight those target sites (if any) for which three-finger proteins have already been designed

iii) whether the program should re-order output target sites according to their relative positions in the input target sequence

iv) whether the program should highlight targetable pairs of 9-bp DNA sites (adjacent, nonoverlapping site pairs separated by n or fewer bases, where n is typically 5, 4, 3, 2 or 1).

Output: A set of potential target sites in the target DNA sequence ranked by score.

If specified, a list of any target sites for which three-finger proteins have already been designed.

If specified, the list of output sites re-ordered according to location in the input sequence

If specified, a list of all targetable pairs of 9-bp DNA sites.

The site selection portion of the program assigns a score to every possible 9-bp sequence in a given target DNA fragment, the score reflecting ease of targetability based on using information from previously designed zinc finger proteins. In evaluating a given 9 base sequence, the program first splits the target into its component subsites, and then consults the scores table to obtain a score for each subsite at its location in the potential target site. Finally, it multiplies the subsite scores to obtain an overall score for the 9-bp target site. For example, using the test sequence 5'AGTGCGCGGTGC3' and the scores table in Table 1, the output sites (5'-3') and scores are

site	subsites	score	
AGTGCGCGG	AGT GCG CGG	1 x 10 x 1	= 10
GTGCGCGGT	GTG CGC GGT	10 x 1 x 10	= 100
TGCGCGGTG	TGC GCG GTG	10 x 10 x 10	= 1000
GCGCGGTGC	GCG CGG TGC	10 x 1 x 8	= 80

In this example, the best target site is 5'TGC GCG GTG3', with a score of 1000. The program also assigns scores to potential targets in the opposite (antisense) strand, but for the sake of simplicity these sites are ignored in this example.

A optional factor, the "enhancement factor for 'D-able' triplets", can be provided to alter the above scoring protocol to account for the context factor - the D-able contact - in evaluating target sites. If this feature is chosen, the program performs the following check when assigning subsite scores:

If a subsite is of the form xxG, then if the adjacent base (on the 3' side) is T or G, then the score of the xxG subsite is multiplied by the enhancement factor, otherwise, the subsite score remains the same.

[If the subsite is of the form xxA, xxC or xxT, its score also remains unchanged.]

For example, if the user inputs an enhancement factor for 'D-able' triplets of 1.25, then the scores above are adjusted as follows:

site	subsites	score
AGTGCGCGGt	AGT GCG CGGt	1x10x(1x1.25) = 12.5 (CGG is D-able)
GTGCGCGGTg	GTG CGC GGTg	10x1x10 (no D-able subsites)
TGCGCGGTGc	TGC GCG GTGc	10x(10x1.25)x10= 1250 (GCG is D-able)
GCGCGGTGC#	GCG CGG TGC#	10x(1x1.25)x8 = 100 (CGG is D-able)

[When using this option, the program considers the identity of base immediately to the 3' side of the target site (in lower case). For the last site, this base is undefined in this example and this is noted by placing the pound sign '#' at this position.]

After assigning scores to all 9-base pair sequences in the target DNA, the program then prints out the top scores, with the number of sites outputted determined by the user.

As specified by the user, the program can also provide:

i. a list of any target sites for which three-finger proteins have already been designed.

ii the list of output sites re-ordered according to location in the input sequence

iii a list of all targetable pairs of 9-bp DNA sites (adjacent, nonoverlapping site pairs separated by five, three or fewer bases).

II. Design of proteins for the chosen target sites

Inputs: Sites from the site-selection portion of the program (or otherwise determined)

The 'ZFPdata table' which contains target sites, amino acid sequences, and reference data for existing high-affinity ZFPs.

An output parameter - supplied by the user - specifying whether the program should restrict its output either:

i to only those proteins (if any) whose target sites are completely identical to sites in the output, or,

ii to only those proteins (if any) whose target sites match output sites at two or more of the three-bp subsites.

Output: In the absence of restrictions (i) or (ii):

For each potential 9-base pair target site, a listing of three sets of ZFPs and their component fingers from the ZFP data table which respectively bind to the three triplet subsites within the target site. For each subsite, the set of ZFPs can be subdivided into two subsets. One subset contains ZFPs and their fingers that bind a triplet at a given position from the corresponding finger position in a parental ZFP. The other subset contains ZFPs and their fingers that bind a triplet at a given position from a noncorresponding position within a parental ZFP. A first finger position (N-C) corresponds to a first triplet position 3'-5'.

The ZFP design portion of the program facilitates the design process by allowing the user to rapidly review all fingers known to bind subsites in a given 9-base target site. target. Given the optimal design target from the above example (5'TGCGCGGTG3'), and the short ZFP data table provided in Table 9, the output (in the absence of restrictions (i) or (ii)) would be as follows:

site score
5'TGCGCGGTG 1.00E+003

ZFPs -- PREVIOUS DESIGN:

ORDERED:

Triplet	3	2	1	F1	F2	F3
[1]	5'TGCGGGGCA			********	********	*ERDHLRT
[3]	5'GGGGCGGGG			********	*RSDELQR	********
[4]	5'GAGTGTGTG			*RKDSLVR	********	********

DISORDERED:

********	*RSDELTR[2](3)	********
********	*RSDERKR[2](1)	********

The 'ordered' output shows that, in the ZFP data table, ther is one instance where the TGC subsite is contacted by a zinc finger in the third triplet of a target site.

The finger in this case is ERDHLRT, and the site is 5'TGCGGGGCA3'. There is also one similar instance for each of the other two subsites - GCG, and GTG. The fingers in these cases are, respectively, RSDELQR and RKDSLVR. This information is used to propose the three finger protein F1-RKDSLVR, F2-RSDELQR, F3-ERDHLRT as a design to bind the target 5'TGCGCGGTG3'.

The 'disordered' output shows that there are two cases in the ZFPdata table in which fingers contact a GCG subsite, but not at the center subsite of a target. Rather, in one case GCG is contacted at the 5' end, and the other the 3' end, and in these cases the finger sequences are RSDELTR and RSDERKR. These are alternate designs for binding GCG in the target site.

Table 1: The scores table

subsite sequence:	subsite score: location in 9 base site: base pairs # 7-9	4-6	1-3	subsite sequence:	subsite score: location in 9 base site: base pairs # 7-9	4-6	1-3
AAA	10	8	8	CAA	8	8	10
AAG	8	8	10	CAG	1	1	1
AAC	1	1	1	CAC	1	1	1
AAT	8	10	10	CAT	1	1	1
AGA	10	8	8	CGA	1	1	1
AGG	1	1	1	CGG	1	1	1
AGC	1	1	1	CGC	1	1	1
AGT	1	1	1	CGT	1	1	1
ACA	8	10	8	CCA	1	1	1
ACG	1	1	1	CCG	1	1	1
ACC	1	1	1	CCC	1	1	1
ACT	1	1	1	CCT	1	1	1
ATA	8	10	8	CTA	1	1	1
ATG	1	1	1	CTG	1	1	1
ATC	1	1	1	CTC	1	1	1
ATT	1	1	1	CTT	1	1	1
GAA	10	10	10	TAA	8	8	10
GAG	10	10	10	TAG	10	10	8
GAC	10	10	8	TAC	10	8	10
GAT	10	10	10	TAT	1	1	1
GGA	10	10	10	TGA	10	10	8

	GGG	10	10	10	TGG	10	10	10
	GGC	10	10	10	TGC	8	10	10
	GGT	10	10	10	TGT	10	10	8
	GCA	10	10	10	TCA	10	8	8
	GCG	10	10	10	TCG	8	10	8
	GCC	10	10	8	TCC	10	8	10
	GCT	10	10	10	TCT	1	1	1
	GTA	10	10	10	TTA	10	10	8
	GTG	10	10	10	TTG	8	10	8
	GTC	10	10	10	TTC	1	1	1
	GTT	10	10	10	TTT	8	10	8

Table 9: Exemplary ZFP data table

#	target site	ZFP sequence F1	F2	F3	reference information
1	TGCGGGGCA	RSADLTR	RSDHLTR	ERDHLRT	SBS design GR-223, Kd: 8 nM
2	GCGTGGGCG	RSDELTR	RSDHLTT	RSDERKR	Zif 268, Kd: 0.04 nM
3	GGGGCGGGG	KTSHLRA	RSDELQR	RSDHLSK	SP1, Kd: 25 nM
4	GAGTGTGTG	RKDSLVR	TSDHLAS	RSDNLTR	SBS design GL-8.3.1, Kd: 32 nM

Other examples of zinc finger proteins, the sequences of their fingers and target sites bound appropriate for inclusion in such a database are discussed in the references cited in the Background Section.

Although the foregoing invention has been described in detail for purposes of clarity of understanding, it will be obvious that certain modifications may be practiced within the scope of the appended claims. All publications and patent documents cited herein are hereby incorporated by reference in their entirety for all purposes to the same extent as if each were so individually denoted.

WHAT IS CLAIMED IS:

Selection of D-able subsites

1. A method of selecting a target site within a target sequence for targeting by a zinc finger protein comprising:

providing a target nucleic acid to be targeted by a zinc finger protein;

outputting a target site within the target nucleic acid comprising 5'NNx aNy bNzc3', wherein

each of (x, a), (y, b) and (z, c) is (N, N) or (G, K);

at least one of (x, a), (y, b) and (z, c) is (G, K). and

N and K are IUPAC-IUB ambiguity codes.

2. The method of claim 1, further comprising selecting a plurality of potential target sites within the target nucleic acid and outputting a subset of the plurality of potential target segments comprising 5'NNx aNy bNzc3', wherein

each of (x, a), (y, b) and (z, c) is (N, N) or (G, K);

at least one of (x, a), (y, b) and (z, c) is (G, K). and

N and K are IUPAC-IUB ambiguity codes.

3. The method of claim 2, wherein the target nucleic acid comprises a target gene.

4. The method of claim 1, wherein at least two of (x, a), (y, b) and (z, c) is (G, K).

5. The method of claim 1, wherein all three of (x, a), (y, b) and (z, c) are (G, K).

6. The method of claim 1, wherein the zinc finger protein comprises three fingers.

7. The method of claim 1, wherein the target site comprises first and second target segments, each comprising 5'NNx aNy bNzc3', and the method further comprises selecting the second target segment.

8. The method of claim 7, wherein in the second segment at least two of (x, a), (y, b) and (z, c) are (G, K).

9. The method of claim 8, wherein in the second segment all three of (x, a), (y, b) and (z, c) are (G, K).

10. The method of claim 10, wherein the first and second target segment are separated by fewer than 5 bases in the target site.

11. The method of claim 10, wherein the first target segment comprises 5'NNN NNN NNG3', the second target segment comprises 5'KNx aNY bNzc3' and there are zero bases separating the first and second target segments in the target site.

12. The method of claim 7, further comprising synthesizing step comprises synthesizing a first zinc finger protein comprising three zinc fingers that respectively bind to the NNx aNy and bNz triplets in the target segment and a second three fingers that respectively bind to the NNx aNy and bNz triplets in the second target segment.

13. The method of claim 1, further comprising synthesizing a zinc finger protein comprising first, second and third fingers that bind to the bNz aNy and NNx triplets respectively.

14. The method of claim 13, wherein each of the first, second and third fingers is selected or designed independently.

15. The method of claim 13, wherein a finger is designed from a database containing designation of zinc finger proteins, subdesignations of finger components, and nucleic acid sequences bound by the zinc finger proteins.

16. The method of claim 13, wherein a finger is selected by screening variants of a zinc finger binding protein for specific binding to the target site to identify a variant that binds to the target site.

17. The method of claim 13, further comprising contacting a sample containing the target nucleic acid with the zinc finger protein, whereby the zinc finger protein binds to the target site revealing the presence of the target nucleic acid or a particular allelic form thereof.

18. The method of claim 13, further comprising contacting a sample containing the target nucleic acid with the zinc finger protein, whereby the zinc finger protein binds to the target site thereby modulating expression of the target nucleic acid

19. The method of claim 1, wherein the target site occurs in a coding region of a gene

20. The method of claim 1, wherein the target site occurs within or proximal to a promoter, enhancer, or transcription start sit

21. The method of claim 1, wherein the target site occurs outside a promoter, regulatory sequence or transcriptional start site within the target nucleic acid.

Selection of Target Sites Using a Correspondence Regim

22. A method for selecting a target site within a polynucleotide for targeting by a zinc finger protein, comprising:

providing a polynucleotide sequence;

selecting a potential target site of within the polynucleotide sequence; the potential target site comprising contiguous first, second and third triplets of bases at first, second and third positions in the potential target site;

determining a plurality of subscores by applying a correspondence regime between triplets and triplet position in a sequence of three contiguous triplets, wherein each triplet has first, second and third corresponding positions, and each combination of triplet and triplet position has a particular subscore

calculating a score for the potential target site by combining subscores for the first, second, and third triplets;

repeating the selecting, determining and calculating steps at least once on a further potential target site comprising first, second and third triplets at first, second and third positions of the further potential target site to determine a further score;

providing output of at least one potential target site with its score.

23. The method of claim 22, wherein output is provided of the potential target site with the highest score.

24. The method of claim 22, wherein output is provided of the n potential target sites with the highest scores, and the method further comprises providing user input of a value for n.

25. The method of claim 22, wherein the subscores are combined by forming the product of the subscores.

26. The method of claim 25, wherein the correspondence regime comprises 64 triplets, each having first, second, and third corresponding positions, and 192 subscores.

27. The method of claim 22, wherein the subscores in the correspondence regime are determined by assigning a first value as the subscore of a subset of triplets and corresponding positions, for each of which there is an existing zinc finger protein that comprising a finger that specifically binds to the triplet from the same position in the existing zinc finger protein as the corresponding position of the triplet in the correspondence regime; assigning a second value as the subscore of a subset of triplets and corresponding positions, for each of which there is an existing zinc finger protein that comprises a finger that specifically binds to the triplet from a different position in the existing zinc finger protein than the corresponding position of the triplet in the correspondence regime; and assigning a third value as the subscore of a subset of triplets and corresponding positions for which there is no known zinc protein comprising a finger that specifically binds to the triplet.

28. The method of claim 22, wherein the correspondence regime is shown in Table 1.

29. The method of claim 22, further comprising combining a context parameter with the subscore of at least one of the first, second and third triplets to give a scaled subscore of the at least one triplet.

30. The method of claim 29, wherein the context parameter is combined with the subscore when the target site comprises a base sequence 5'NNGK3', wherein NNG is the at least one triplet.

31. The method of claim 22, further comprising combining a context parameter that is combined with the score of a potential target site to give a scaled score

32. The method of claim 31, wherein the context parameter is combined with the score when a potential target site comprises 5'NNx aNy bNzc3', wherein

wherein each of (x, a), (y, b) and (z, c) is (N, N) or (G, K);

at least one of (x, a), (y, b) and (z, c) is (G, K). and

N and K are IUPAC-IUB ambiguity codes.

33. The method of claim 32, wherein a first context parameter is combined with the score if one of (x, a), (y, b) and (z, c) is (G, K), and a second context parameter is combined with the score if two of (x, a), (y, b) and (z, c) are (G, K), and a third context parameter is input if three of (x, a), (y, b) and (z, c) are (G, K)

34. The method of claim 22, wherein output is provided of at least a nonoverlapping pair of potential target sites and their scores, the members of the pair being separated by five or fewer bases in the polynucleotide.

Design of ZFPs using a Database

35. A method of producing a zinc finger protein comprising:

(a) providing a database comprising designations for a plurality of zinc finger proteins, each protein comprising at least first, second and third fingers, and subdesignations for each of the three fingers of each of the zinc finger proteins;

a corresponding nucleic acid sequence for each zinc finger protein, each sequence comprising at least first, second and third triplets specifically bound by the at least first, second and third fingers respectively in each zinc finger protein, the first, second and third triplets being arranged in the nucleic acid sequence (3'-5') in the same respective order as the first, second and third fingers are arranged in the zinc finger protein (N-terminal to C-terminal);

(b) providing a target site for design of a zinc finger protein, the target site comprising continuous first, second and third triplets in a 3'-5' order,

(c) for the first, second and third triplet in the target site, identifying first, second and third sets of zinc finger protein(s) in the database, the first set comprising zinc finger protein(s) comprising a finger specifically binding to the first triplet in the target site, the second set comprising zinc finger protein(s) comprising a finger specifically binding to the second triplet in the target site, the third set comprising zinc finger protein(s) comprising a finger specifically binding to the third triplet in the target site;

(d) outputting designations and subdesignations of the zinc finger proteins in the first, second, and third sets identified in step (c).

36. The method of claim 35, further comprising:

(e) producing a zinc finger protein that binds to the target site comprising a first finger from a zinc finger protein from the first set, a second finger from a zinc finger protein from the second set, and a third finger from a zinc finger protein from the third set.

37. The method of claim 36 further comprising identifying subsets of the first, second and third sets, the subset of the first set comprising zinc finger protein(s) comprising a finger that specifically binds to the first triplet in the target site from the first finger position of a zinc finger protein in the database; the subset of the second set comprising zinc finger protein(s) comprising a finger that specifically binds to the second triplet in the target site from the second finger position in a zinc finger protein in the database; the subset of the third set comprising zinc finger protein(s) comprising a finger

that specifically binds to the third triplet in the target site from a third finger position in a zinc finger protein in the database;

wherein

the outputting step comprising outputting designations and subdesignations of the subset of the first, second and third sets; and

the producing step comprising producing a zinc finger protein comprising a first finger from the first subset, a second finger from the second subset, and a third finger from the third subset.

38. The method of claim 37, wherein the outputting comprises outputting the designations and subdesignations of the subsets of the first, second and third sets, and the first, second and third sets minus their respective subsets.

39. The method of claim 38, wherein each of the subsets is a null set.

40. The method of claim 35, wherein the target site is provided by user input.

41. The method of claim 35 wherein the target site is provided by the method of claim 1 or claim 22.

42. A method of producing a zinc finger protein comprising:

(a) providing a database comprising

designations for a plurality of zinc finger proteins, each protein comprising at least first, and second fingers,

subdesignations for each of the fingers of each of the zinc finger proteins; and

a corresponding nucleic acid sequence for each zinc finger protein, each sequence comprising first and second triplets specifically bound by the first and second fingers respectively, the triplets being arranged in the nucleic acid sequence (3'-5') in the same respective order as the first and second and fingers are arranged in the zinc finger protein (N-terminal to C-terminal);

(b) providing a target site for design of a zinc finger protein, the target site comprising contiguous first, and second triplets ordered 3'5' in the target site;

(c) for the first and second triplet in the target site, identifying first and second sets of zinc finger protein(s) in the database, the first set comprising zinc finger protein(s) comprising a finger specifically binding to the first triplet in the target site, the second set comprising zinc finger protein comprising a finger specifically binding to the second triplet in the target site;

(d) outputting designations and subdesignations of the zinc finger proteins in the first, and second sets identified in step (c).

43 A method of producing a zinc finger protein comprising:

(a) providing a database comprising:

designations for a plurality of zinc finger proteins, each protein comprising at least first, and second fingers;

subdesignations for each of the fingers of each of the zinc finger proteins; and

a corresponding nucleic acid sequence for each zinc finger protein, each sequence comprising first, and second triplets specifically bound by the first and second fingers respectively, the triplets being arranged in the nucleic acid sequence (3'-5') in the same respective order as the first and second and fingers are arranged in the zinc finger protein (N-terminal to C-terminal);

(b) providing a target site for design of a zinc finger protein, the target site comprising contiguous first, second and third triplets ordered 3'5' in the target site;

(c) for the first and third triplet in the target site, identifying first and second sets of zinc finger protein(s) in the database, the first set comprising zinc finger protein(s) comprising a finger specifically binding to the first triplet in the target site, the second set comprising zinc finger protein comprising a finger specifically binding to the third triplet in the target site;

(d) outputting designations and subdesignations of the zinc finger proteins in the first, and second sets identified in step (c).

44 A computer program product for selecting a target sequence within a polynucleotide for targeting by a zinc finger protein, comprising:

(a) code for providing a polynucleotide s quence;

(b) code for selecting a potential target site within the polynucleotide sequence; the potential target site comprising first, second and third triplets of bases at first, second and third positions in the potential target site;

(c) code for calculating a score for the potential target site from a combination of subscores for the first, second, and third triplets, the subscores being obtained from a correspondence regime between triplets and triplet position, wherein each triplet has first, second and third corresponding positions, and each corresponding triplet and position has a particular subscore;

(d) code for repeating steps (b) and (c) at least once on a further potential target site comprising first, second and third triplets at first, second and third positions of the further potential target site to determine a further score;

(e) code for providing output of at least one of the potential target site with its score

(f) a computer readable storage medium for holding the codes

45. The computer program product of claim 44, further comprising code for combining a context parameter with a subscore.

46. A system for selecting a target sequence within a polynucleotide for targeting by a zinc finger protein, comprising:

(a) a memory;

(b) a system bus;

(c) a processor operatively disposed to:

(1) provide or receive a polynucleotide sequence;

(2) select a potential target site within the polynucleotide sequence; the potential target site comprising first, second and third triplets of bases at first, second and third positions in the potential target site;

(3) calculate a score for the potential target site from a combination of subscores for the first, second, and third triplets, the subscores being obtained from a correspondence regime between triplets and triplet position, wherein each triplet has first, second and third corresponding positions, and each corresponding triplet and position has a particular subscore;

(4) repeat steps (2) and (3) at least once on a further potential target site comprising first, second and third triplets at first, second and third positions of the further potential target site to determine a further score;

(5) provide output of at least one of the potential target site with its score.

47. The system of claim 46, wherein the processor is further operatively disposed to combine a context parameter with a subscore.

48. A computer program product for designing a zinc finger protein comprising:

(a) code for providing a database comprising

designations for a plurality of zinc finger proteins, each protein comprising at least first, second and third fingers ,

subdesignations for each of the three fingers of each of the zinc finger proteins;

a corresponding nucleic acid sequence for each zinc finger protein, each sequence comprising at least first, second and third triplets specifically bound by the at least first, second and third fingers respectively in each zinc finger protein, the first, second and third triplets being arranged in the nucleic acid sequence (3'-5') in the same respective order as the first, second and third fingers are arranged in the zinc finger protein (N-terminus to C-terminus);

(b) code for providing a target site for design of a zinc finger protein, the target site comprising at least first, second and third triplets,

(c) for the first, second and third triplet in the target site, code for identifying first, second and third sets of zinc finger protein(s) in the database, the first set comprising zinc finger protein(s) comprising a finger specifically binding to the first triplet in the target site, the second set comprising a finger specifically binding to the second triplet in the target site, the third set comprising a finger specifically binding to the third triplet in the target site;

(d) code for outputting designations and subdesignations of the zinc finger proteins in the first, second, and third sets identified in step (c).

(e) a compute readable storage medium for holding the codes.

49. A system for designing a zinc finger protein comprising:

(a) a memory;

(b) a system bus;

(c) a processor operatively disposed to:

(1) provide a database comprising designations for a plurality of zinc finger proteins, each protein comprising at least first, second and third fingers, subdesignations for each of the three fingers of each of the zinc finger proteins;

a corresponding nucleic acid sequence for each zinc finger protein, each sequence comprising at least first, second and third triplets specifically bound by the at least first, second and third fingers respectively in each zinc finger protein, the first, second and third triplets being arranged in the nucleic acid sequence (3'-5')in the same respective order as the first, second and third fingers are arranged in the zinc finger protein (N-terminus to C-terminus);

(2) provide or be provided with a target site for design of a zinc finger protein, the target site comprising at least first, second and third triplets,

(3) for the first, second and third triplet in the target site, identifying first, second and third sets of zinc finger protein(s) in the database, the first set comprising zinc finger protein(s) comprising a finger specifically binding to the first triplet in the target site, the second set comprising a finger specifically binding to the second triplet in the target site, the third set comprising a finger specifically binding to the third triplet in the target site;

(4) output designations and subdesignations of the zinc finger proteins in the first, second, and third sets identified in step (3).

50. A computer program product for selecting a target site within a target sequence for targeting by a zinc finger protein comprising:

code for providing a target nucleic acid to be targeted by a zinc finger protein;

code for outputting a target site within the target nucleic acid comprising 5'NNx aNy bNzc3', wherein

wherein each of (x, a), (y, b) and (z, c) is (N, N) or (G, K);

at least one of (x, a), (y, b) and (z, c) is (G, K). and

N and K are IUPAC-IUB ambiguity codes;

and a computer readable storage medium for holding the codes.

51. A system for selecting a target site within a target sequence for targeting by a zinc finger protein comprising:

(a) a memory;

(b) a system bus;

(c) a processor operatively disposed to:

provide a target nucleic acid to be targeted by a zinc finger protein;

output a target site within the target nucleic acid comprising 5'NNx aNy bNzc3', wherein

wherein each of (x, a), (y, b) and (z, c) is (N, N) or (G, K);

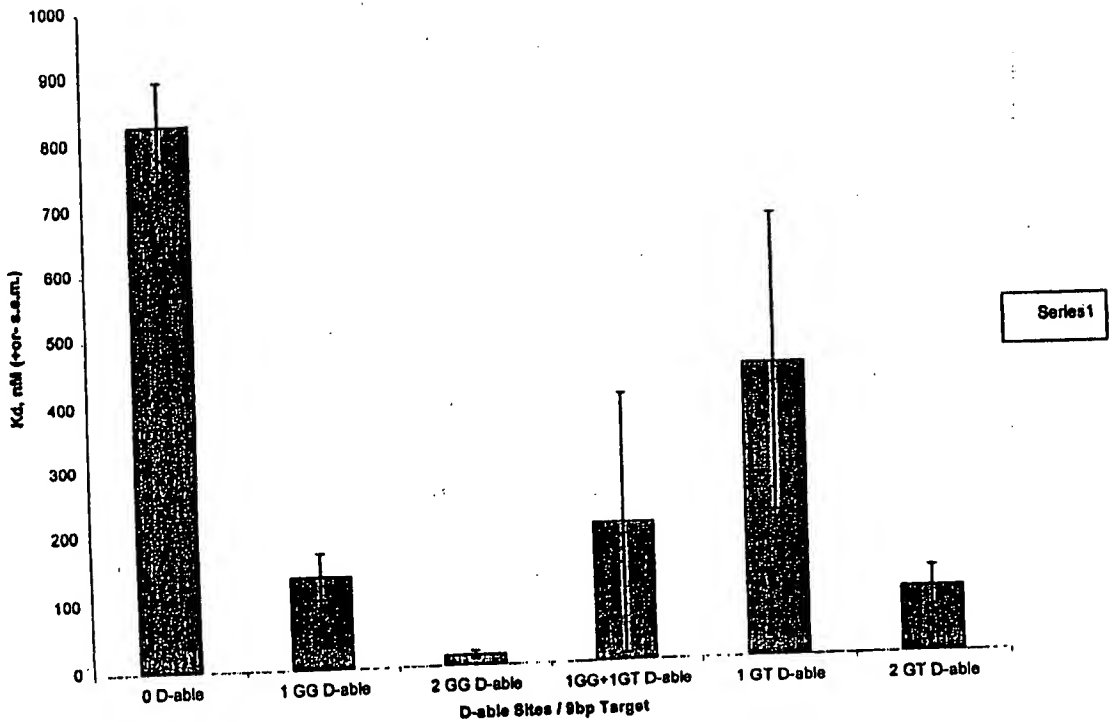
at least one of (x, a), (y, b) and (z, c) is (G, K) and

N and K are IUPAC-IUB ambiguity codes.

Fig. 1

Kd of Designed ZFPs vs. D-able Sites / Target:

Affinity is a function of number and type of D-able site



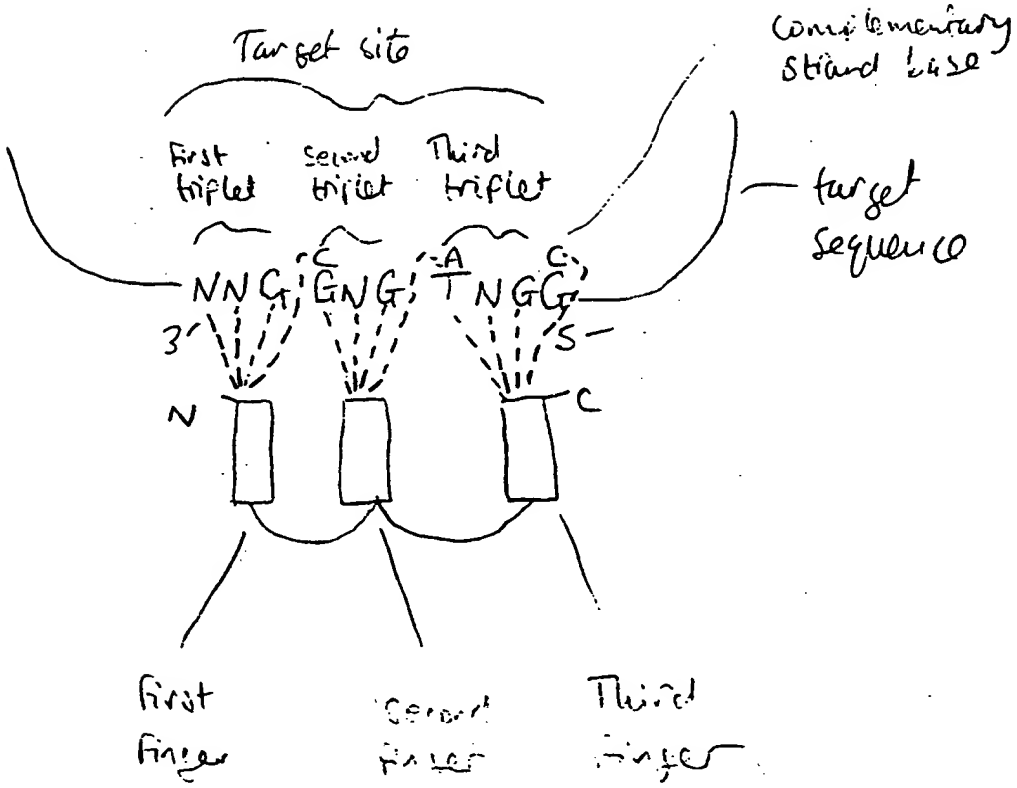

Target site
Complementary Strand base
First triplet
Second triplet
Third triplet
target sequence
NNG
GNG
TNG
C
A
C
3'
5'
N
C
First finger
Second finger
Third finger

Fig. 2

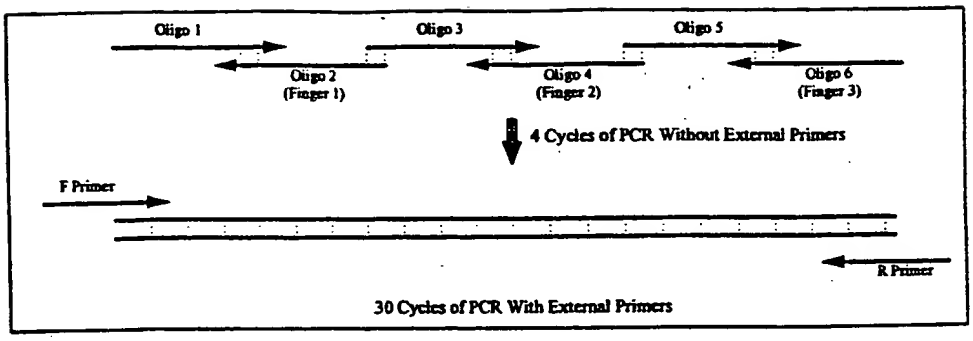


*Figure 3. PCR amplification scheme for production of ZFP-encoding synthetic genes.

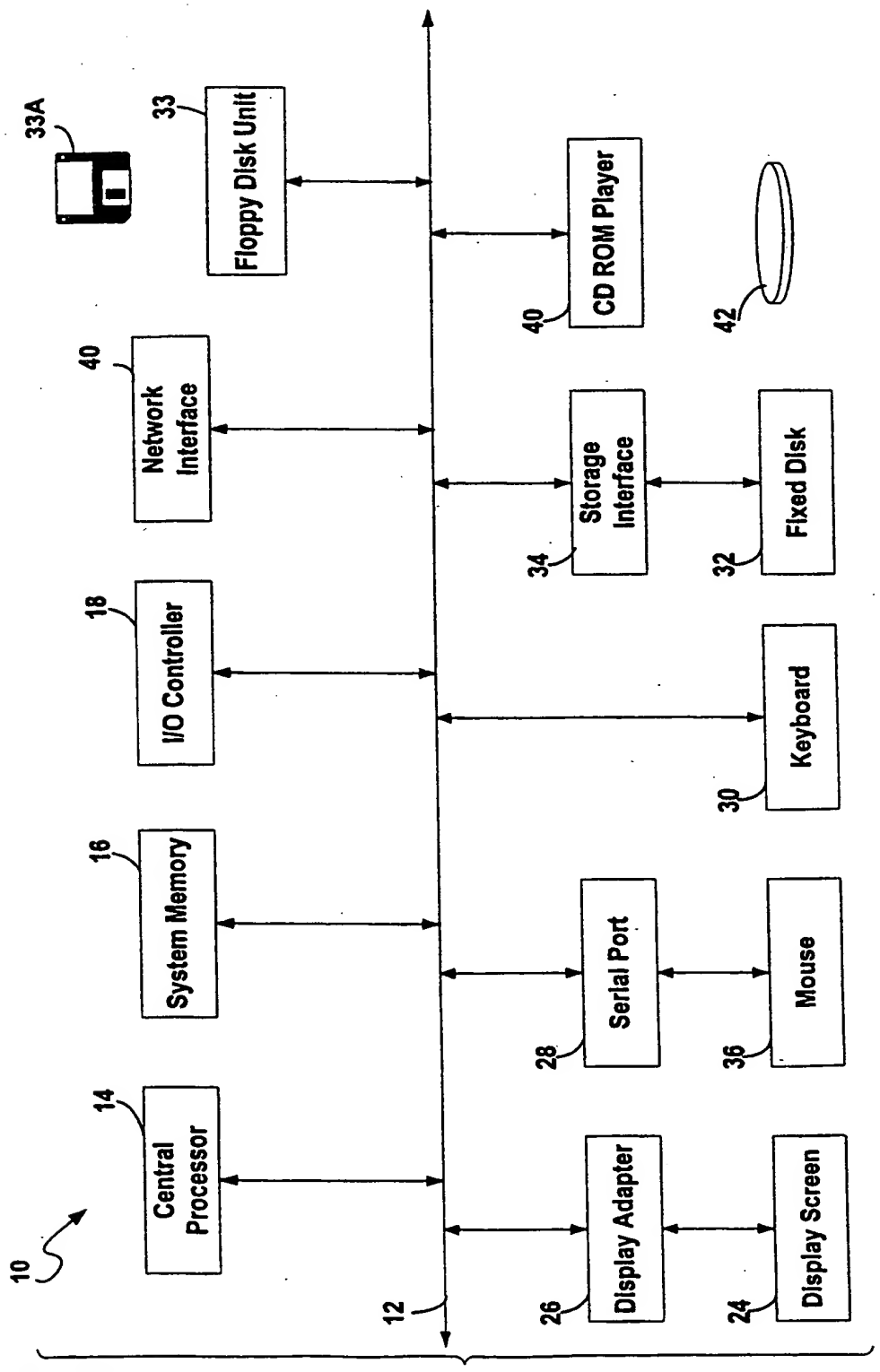


Fig. 4

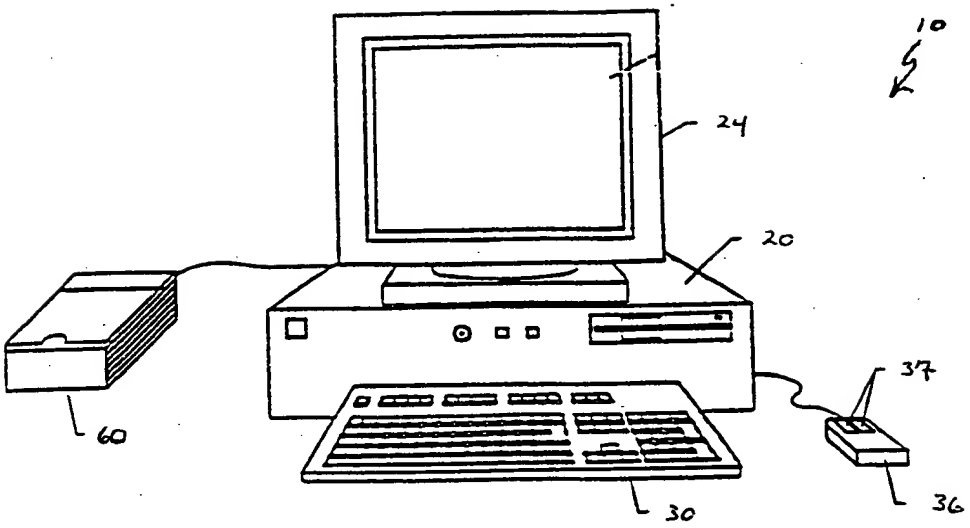


Fig. 5

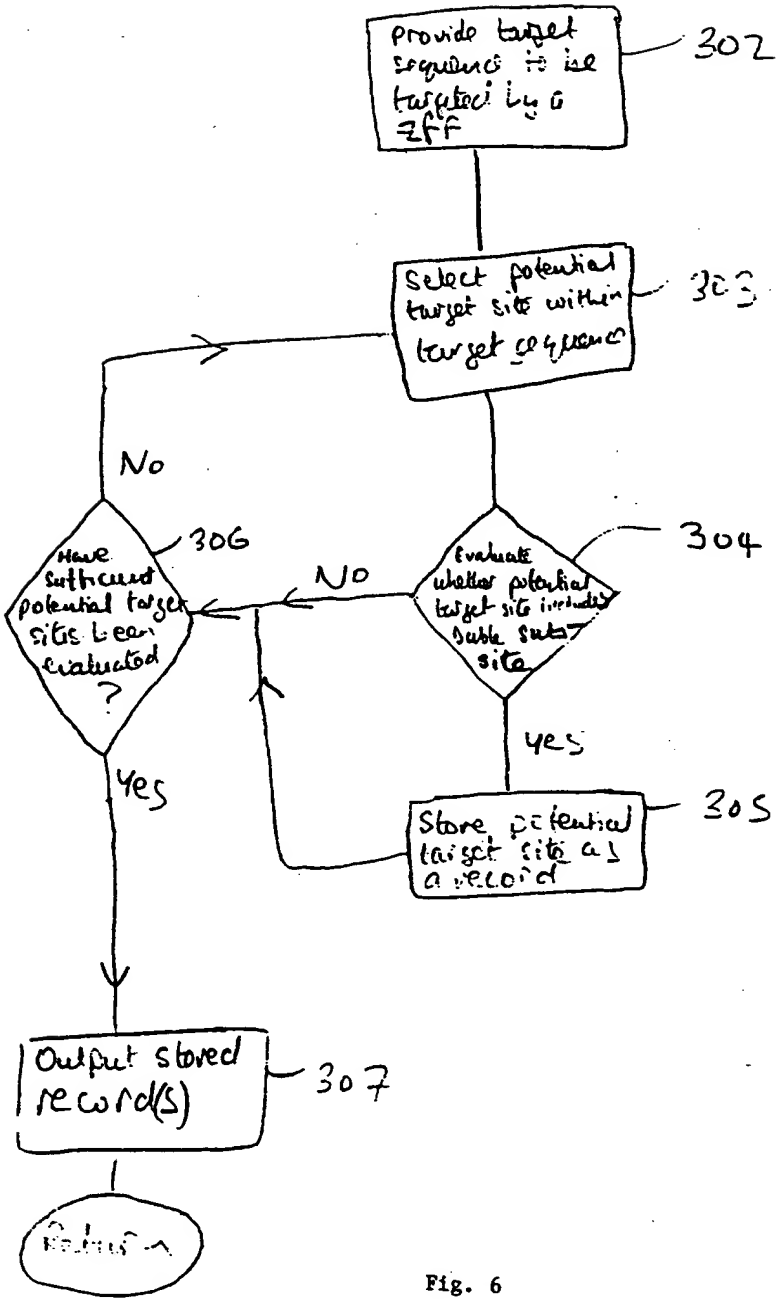

Provide target sequence to be targeted by a ZFP
302
Select potential target site within target sequence
303
Evaluate whether potential target site includes stable subsite
304
No
Have sufficient potential target sites been evaluated ?
306
No
Yes
Store potential target site as a record
305
Yes
Output stored record(s)
307

Fig. 6

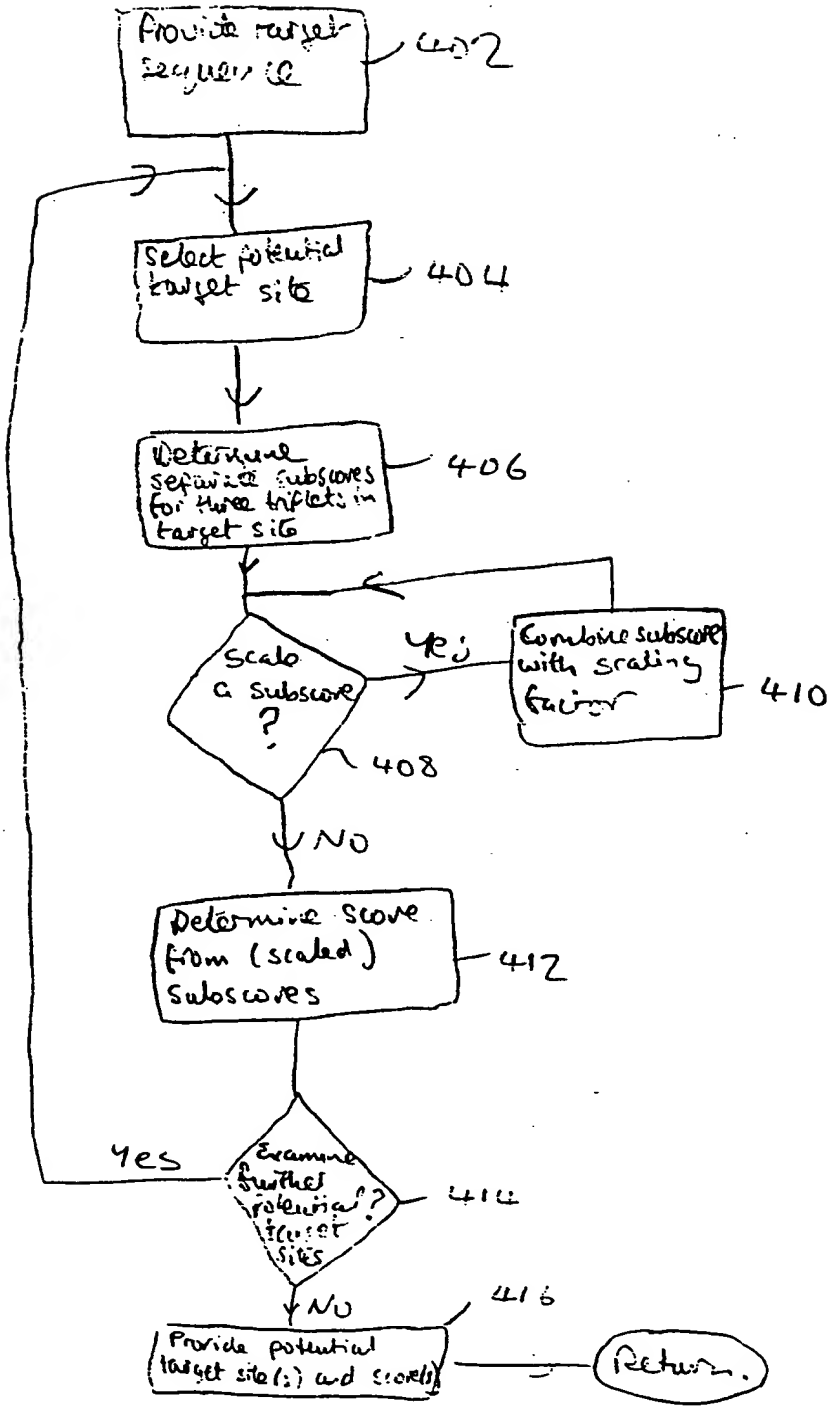

Provide target sequence
402
Select potential target site
404
Determine separate subscores for three triplets in target site
406
Scale a subscore ?
408
Yes
Combine subscore with scaling factor
410
No
Determine score from (scaled) subscores
412
Examine further potential target sites ?
414
Yes
No
416
Provide potential target site(s) and score(s)
Return.

Fig. 7A

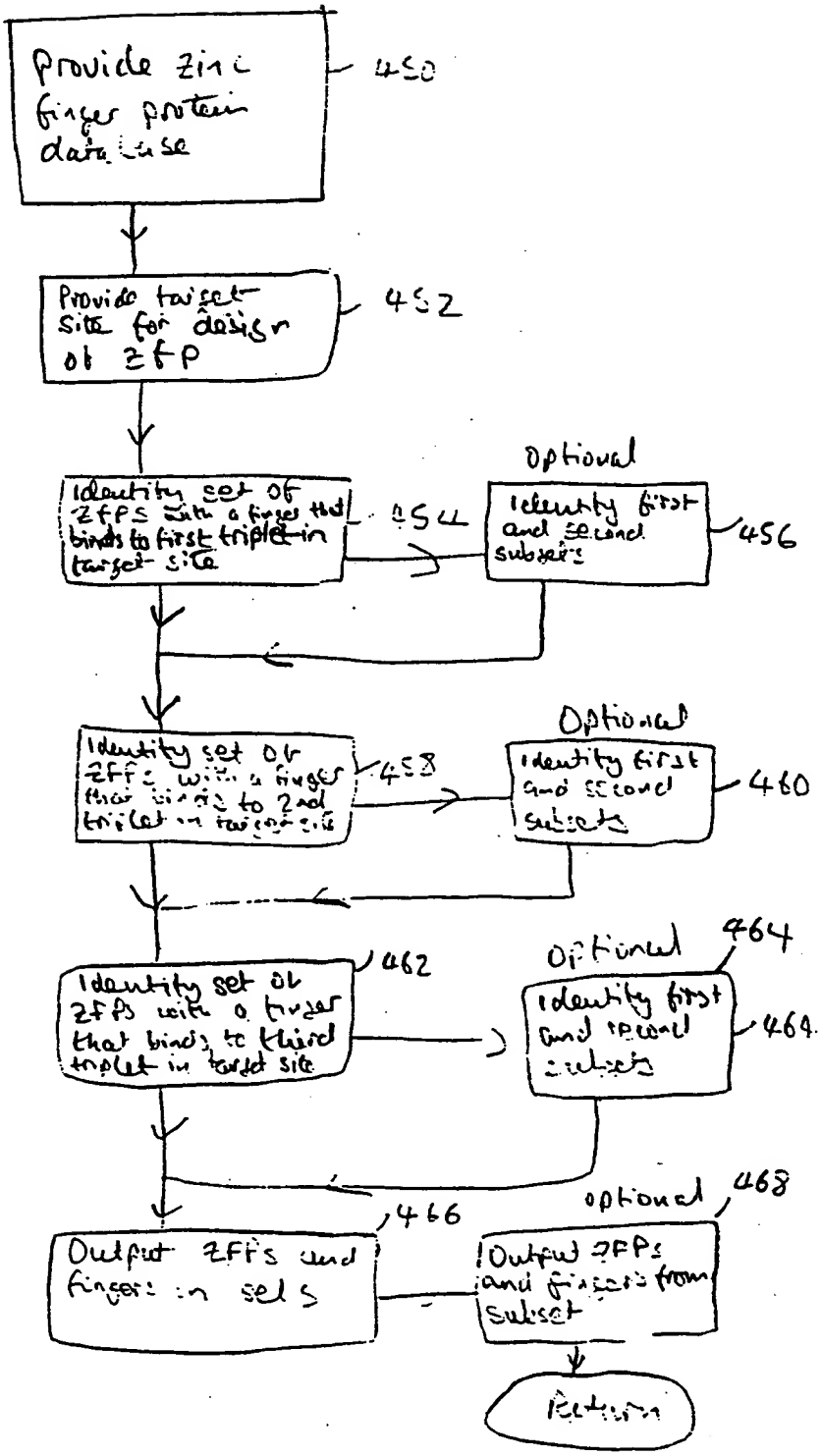

Provide zinc finger protein database
450
Provide target site for design of ZFP
452
Identify set of ZFPs with a finger that binds to first triplet in target site
454
Optional
Identify first and second subsets
456
Identify set of ZFPs with a finger that binds to 2nd triplet in target site
458
Optional
Identify first and second subsets
460
Identify set of ZFPs with a finger that binds to third triplet in target site
462
Optional
Identify first and second subsets
464
464
Output ZFPs and fingers in sets
466
optional
Output ZFPs and fingers from subset
468
Return

Fig. 7B

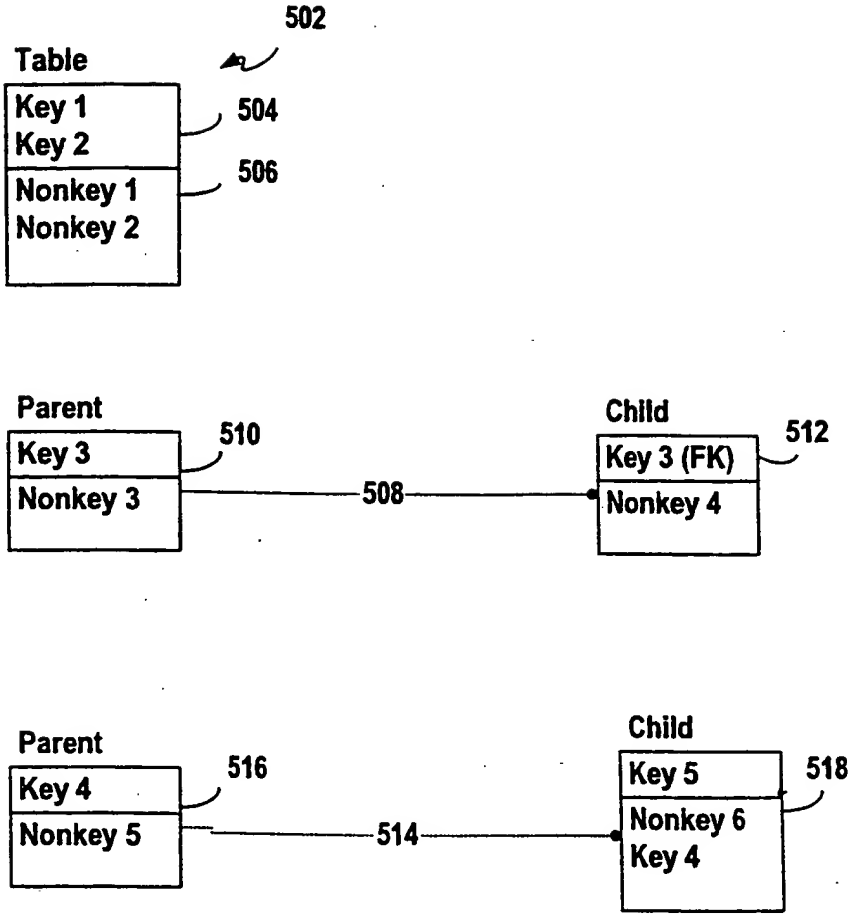

502
Table
Key 1
Key 2
504
Nonkey 1
Nonkey 2
506
Parent
Key 3
Nonkey 3
510
508
Child
Key 3 (FK)
Nonkey 4
512
Parent
Key 4
Nonkey 5
516
514
Child
Key 5
Nonkey 6
Key 4
518

Fig. 8A

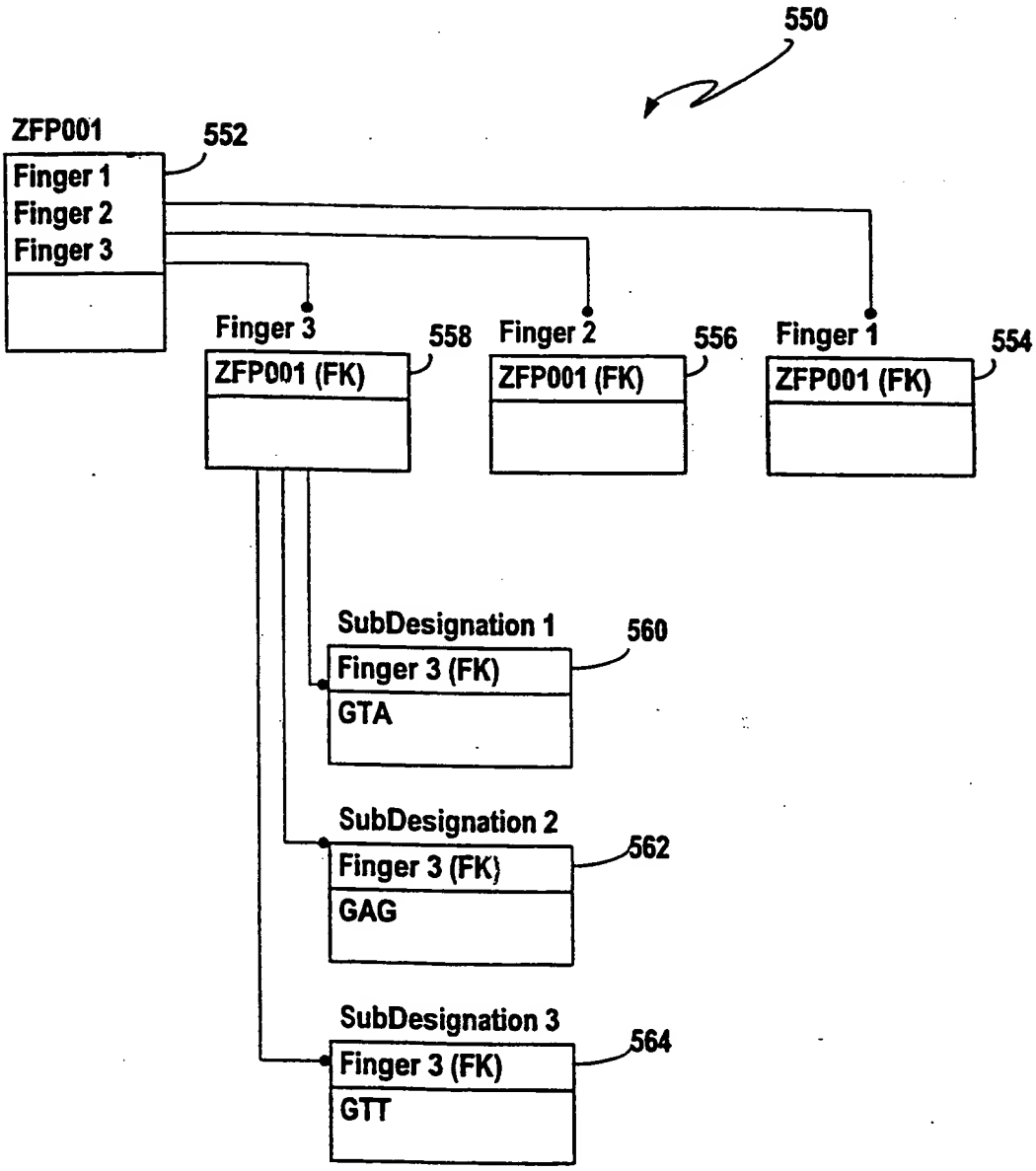


Fig. 8B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/00388

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) :C12Q 1/68
US CL :435/6
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST, STN FILES: MEDLINE, CAPLUS, LIFESCI, EMABASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A,P	US 6,013,453 A (CHOO et al) 11 January 2000, entire document.	1-51
A	US 5,789,538 A (REBAR et al) 04 August 1998, entire document.	1-51

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents.	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 20 MARCH 2000	Date of mailing of the international search report 18 APR 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JANELL TAYLOR JOYCE BRIDGERS PARALEGAL SPECIALIST CHEMICAL MATRIX Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998) *